

## Editorial

Liebe Leserinnen und Leser,

vor Ihnen liegt nunmehr die bereits zwölfte Ausgabe des E-Journals **Anwendungen und Konzepte in der Wirtschaftsinformatik (AKWI)**.

Diese Ausgabe besteht aus sechs wissenschaftlichen Beiträgen, zehn Beiträgen über Anwendungen in Unternehmen, zwei Analysen von Trends und schließlich einen Literaturüberblick über Nachhaltigkeit in der Produktionsplanung.

Die Publikationen über Forschungsarbeiten dieser Ausgabe verwenden alle Simulation und Optimierung. Genauso wie Künstliche Intelligenz profitiert diese Kombination von immer preiswerterer Rechenleistung. Während bei Neuronale Netze methodisch die Konzepte aus den 80-er und 90-er Jahre des letzten Jahrhunderts verwenden, ist die methodische Weiterentwicklung der simultanen Verwendung von Simulation und Optimierung eine notwendige Voraussetzung für ihren Einsatz in der Forschung. Zugleich demonstrieren diese Forschungsarbeiten, dass mit Optimierungsmodellen große Probleme beschrieben und generisch gelöst werden können. Beides ist sehr interessant für die industrielle Praxis, da das aufwändige Suchen nach einer sehr guten (oder gar optimalen) Lösung und ihrer noch aufwendigeren Implementierung entfällt. Zwei Arbeiten lösen „Routing“-Probleme und eines ermittelt einen Produktionsplan für die Herstellung von Gummi-Handschuhen, die sehr wichtig für die thailändische Industrie sind. Die anderen drei Forschungsarbeiten stammen aus der anwendungsnahen Forschung, sei es im Bereich des Systementwurfs via SOA, der Simulation einer prozessorientierten Fertigung zur Ammoniaksynthese und der Erzeugung eines digitalen Zwillinges für den Antriebsstrang eines Fahrzeugs.

Die zehn Artikel über Anwendungen in Unternehmen wurden von Wirtschaftsinformatikern publiziert. Wie in den letzten Jahren auch, werden Probleme wie Liquiditätsplanung und Metadatenmanagement durchaus mit neueren Techniken wie Cloud Computing und zur besseren virtuellen Inbetriebnahme gelöst. Bezogen auf die effiziente und effektive Nutzung von neuen Technologien bleibt Standardanwendungssoftware – etwa von SAP – sehr relevant. Zwei Beispiele sind der in dieser Ausgabe beschriebene Einsatz von Robotic Process Automation und die Nutzung von Social-Media-Daten. Die anderen Artikel unterstützen die Digitalisierung auf unterschiedlichen Ebenen, nämlich durch eine Mobilitätsplattform, digitale Produkt-Service-Systeme, Natural Language Processing sowie Chatbots.

Die Trend-Themen ITIL4 und DevOps werden gegenübergestellt und zur Smart Factory werden die Risiken durch die derzeit bzw. in der nahen Zukunft eingesetzten Technologien aufgezeigt. Weiterhin bedeutungsvoll ist die Berücksichtigung von Themen der Nachhaltigkeit unter anderem in Form des Verbrauchs von Rohstoffen in der Produktion. Eine Erhebung der wissenschaftlichen Publikationen dazu wird vorgestellt und das Ergebnis über eine Homepage kostenlos zur Verfügung gestellt.

Dieses breite Spektrum an Themen belegt, dass Wirtschaftsinformatiker einen signifikanten Beitrag in vielen Unternehmensbereichen leisten. Gegenüber den letzten Jahren kann festgestellt werden, dass die Arbeiten sich weiter von reinen SAP-ERP-Themen in Richtung neue Technologien und Digitalisierung als Ganzes entwickeln.

Über Ihr Interesse an der Zeitschrift freuen wir uns und wünschen Ihnen Freude bei der Lektüre.

Regensburg, Fulda, Wildau und Luzern, im Dezember 2020 – Bleiben Sie gesund!

Frank Herrmann, Norbert Ketterer, Konrad Marfurt und Christian Müller



Christian Müller



Konrad Marfurt



Norbert Ketterer



Frank Herrmann

# AN APPROACH TO CREATING A SIMPLE DIGITAL TWIN FOR OPTIMIZING A SMALL ELECTRIC CONCEPT VEHICLE DRIVETRAIN

Tamás Dóka

Péter Horák, PhD

Department of Machine and Product Design

Faculty of Mechanical Engineering

Budapest University of Technology and Economics

1111, Műegyetem rkp. 3, Budapest, Hungary

Email: doka.tamas@gt3.bme.hu

## KEYWORDS

3D Model-Based Simulation; Digital Twin; Electric Vehicle Drivetrain

## ABSTRACT

Since modeling and simulation are integral tools in engineering, the question is not *if* they should be used in a design process, but rather *how* they should be used to deliver the best solutions. The objective of this paper is to outline an approach to creating a simple Digital Twin for a small electric vehicle drivetrain utilizing only parametric 3D CAD models, widely used simulation tools and some programming libraries. First, the concept of the Digital Twin, its benefits, then the possibilities of using Generative Design are briefly introduced, afterwards electric vehicles' advantages are reviewed. In an example project the properties and opportunities of the 3D CAD- and simulation models are demonstrated. Finally, future improvements and automated optimization opportunities are discussed.

## INTRODUCTION

Before today's modeling and simulation technologies have been emerged, designing a system and ensure its proper behavior was expensive and time-consuming. The only way to test a system in operation was to build it physically and to subject it to effects and impacts that the designers thought would be necessary (Grieves and Vickers 2017). In the second half of the 20th-century, Computer-Aided Design (CAD) software solutions enabled to create different variations of a system relatively easy. Finally, with a Digital Twin, in theory, we could analyze any product's behavior in different environments without having the physical representation itself.

With the recently available computational power and cloud-based services, the use of complex simulations and detailed virtual prototypes are no longer privileges of the biggest companies, but they can be created and also run on personal computers. In this paper, an example is shown how a simple Digital Twin of a product

can be set up and prepared to be used for optimization, using only open-source or student license software solutions. The subject of this example is a small electric concept car's drivetrain; therefore, the properties of a battery electric vehicle are briefly introduced.

## *Digital Twins*

The term Digital Twin (DT) has many slightly different definitions which are slowly changing through the years. "The Digital Twin is a set of virtual information constructs that fully describes a potential or actual physical manufactured product from the micro atomic level to the macro geometrical level. At its optimum, any information that could be obtained from inspecting a physical manufactured product can be obtained from its Digital Twin." (Grieves and Vickers 2017). In that way, a DT could exist without physical representation and could serve as a substitute for real prototypes.

In applications like aerospace or automotive industries, any modification to the product could generate unforeseen consequences to the whole system, so these effects should be adequately modeled and simulated before applying them to the system in operation (Goossens 2017). Not long ago, Digital Twins have been developed in a bottom-up philosophy after the real system was specified. The Digital Twin concept was used to create a virtual environment where a detailed simulation model is running. Based on that model and sensor information from the real physical world, the behavior of the real-world twin could be estimated. This approach has many case studies where the existing system is monitored, the Digital Twin could easily detect occurring problems and practical solutions could be calculated to solve these issues.

Traditionally in the conceptual phase, possible design alternatives were explored by engineers, which required a lot of experience and time. CAD models and simulations were only involved in the process after the design space was narrow enough to analyze, validate, and fabricate the design (Khan and Awan 2018). A detailed model for a product or process is not only capable of examining the system based on real-life data, but it can be leveraged in the concept design phase to

define, test, and evaluate different variants of the system. In this method, the virtual model is not only used for diagnosis but to find out which version of the system should be built in the first place. It is more common today to utilize methods that offer a more standardized description of the models (Rodič 2017), allowing to use optimization algorithms acting directly on the digital model by modifying its parameters and comparing the simulation results. The Digital Twin technology, combined with novel optimization algorithms using artificial intelligence, could generate feasible systems that not only correspond with the requirements but are optimal in the prescribed aspects.

### *Generative Design*

It is usually hard to say how a specific parameter will affect the whole system without knowing the system itself. In many fields during the concept design phase, even experts are using best practices to set up the initial boundaries of the product. This method could lead to sub-optimal solutions even if very precise optimization takes place in later stages.

Generative Design systems are using parametric design, optimization and simulation techniques, which allow engineers to iterate through a large number of design alternatives. Taking a problem definition as input Generative Design systems could create a set of optimal solutions for the given problem (Khan and Awan 2018). An example is shown in Figure 1.

Commercially available Generative Design systems are promising tools for creating complex mechanical parts for a given load- and constraint set using artificial intelligence and topology optimization. This method assumes that the surroundings and functions of the element are known. However, the constraints and loads acting on the part usually depend on other members of the system, so these generated solutions are only suitable for situations where all other components of the system remain the same. Still, the idea that engineers should only carefully define the functions and coarse boundaries of a system and artificial intelligence could do the rest of the work could create previously unimaginable new inventions.

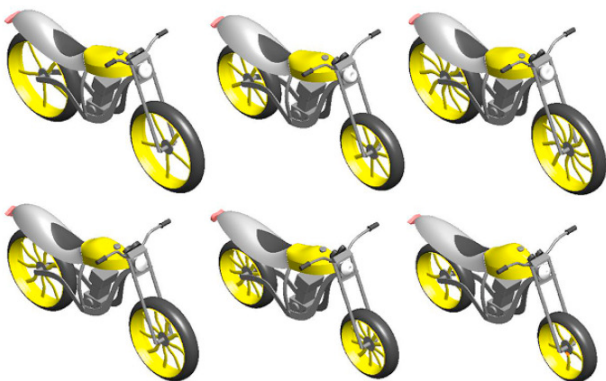


Fig. 1. Generated Design Variations of a Motorcycle (Khan and Awan 2018)

Generative Design combined with Digital Twin technology could allow the system-level optimization where components are heavily co-dependent, such as in electric vehicles.

### *Small Battery Electric Vehicle Concept*

As the present trend suggests, electric vehicles are likely to replace internal combustion engine (ICE) vehicles in the near future (Un-Noor et al. 2017). This trend could be explained by electric vehicles (EVs) being more environmentally friendly, quiet, easy to operate, require less money for fuel and also, they provide instant torque from the startup. EVs are the unquestionably better choice for urban transport, but for longer journeys, two more factors come into question: power and range. Providing a bigger range needs more batteries; therefore, the overall mass of the vehicle grows; thus, the power consumption increases, limiting the achievable range of the vehicle. Finding the right size and arrangement of the electric powertrain components is not as evident as it is in ICE vehicles, because power can be transmitted through electrical wires, enabling to create very different configurations. Moreover, most of the components' parameters are determined by other parts (Un-Noor et al. 2017); thus, optimization becomes even more complicated.

In this paper, an approach for creating a Digital Twin is demonstrated on a small electric vehicle concept. The powertrain of an electric vehicle consists of an electrical and a mechanical subsystem; thus, it is necessary to accurately model and analyze them together to get a real insight into the vehicle's dynamic behavior (Park et al. 2014).

### **APPROACH TO CREATING A SIMPLE DIGITAL TWIN**

As mentioned before, a Digital Twin is basically a set of information about an entity permitting to analyze it from different aspects accurately in a virtual environment. Although every area (mechanical engineering, electrical engineering, etc.) deals with the same product, each of these areas approach the parts that make up these components in a different way (Grieves and Vickers 2017). At present, computers with relatively high computing capacity are affordable, and many modeling, simulating, and optimizing tools and programs are available. Thus the opportunity of using Digital Twins is at hand for smaller businesses and smaller projects too. The design process could be even more improved if we could integrate the available software solutions which we are using, allowing the data exchange between them. Usually, the connection within these software products is not provided, so technically, it requires effort to implement such complex systems. The key to creating a consistent Digital Twin is to persist a homogeneous perspective of the information across functional boundaries. This can be realized by having an application that controls and manages data between different platforms and areas, as shown in Figure 2.

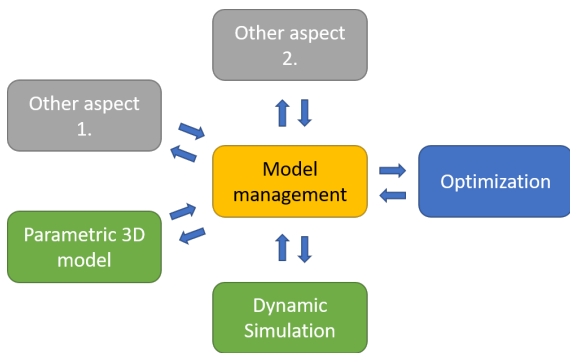


Fig. 2. Block Diagram of the Digital Twin Environment

In the example project, the vehicle was approximately modeled, focusing more on its drivetrain. A detailed, skeleton based top-down 3D model of the drive module (electric motor, fixed gear speed reducer and differential) was modeled and built (see in Figure 3) for further measurements, parameter identification and testing. A simple dynamic simulation was created, using parameters from the CAD models to analyze the vehicle's longitudinal behavior and to provide a basis for optimization in the future.

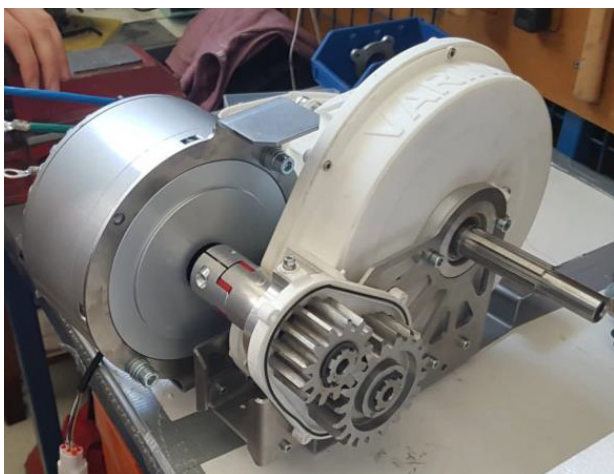


Fig. 3. Drive Module of the Small Electric Vehicle Concept

### Parametric 3D model

A Computer-Aided Design (CAD) model is typically used to visualize the entity, how it will look like in its physical form. Almost every product design starts with an approximate 3D model after the main concept was laid out. These 3D models represent the product's mechanical and physical properties, such as total mass, material properties, or geometric boundaries. Today's advanced CAD systems are capable of performing many different tasks, such as FEM analysis, topology optimization, motion- and dynamic simulations. These integrated CAD systems provide pervasive solutions, but dynamic simulations usually do not require the detailed 3D representation of a product to deliver the desired results or, on the contrary only the aggregated properties of a body (mass, moment of inertia, center of gravity) should be taken into consider-

ation to avoid long computational time.

In the project the vehicle was modeled in PTC Creo 4.0, because it is a high-level CAD system equipped with a wide variety of embedded modules, which can be automated. Make use of the skeleton-based top-down approach, all of the key parameters could be modified, and the assembly regenerated through the managing algorithm.

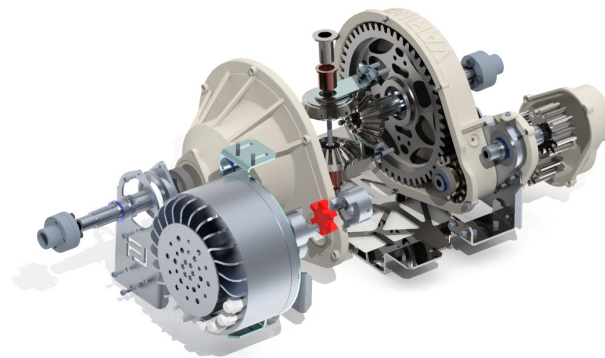


Fig. 4. Exploded View of the Drive Module

### Simulation model

Using only aggregated parameters from the 3D CAD models (not the actual 3D mesh) enables us to simulate the system efficiently. Even without a detailed 3D model, simulation can be processed with approximate parameters for estimations. For example, in the early stages, the total mass of the vehicle is unknown, but the drive module could be tested with different scenarios. Later, when the actual vehicle model is available, approximate parameters could be replaced with precise ones. In the simulation model, a simplified, single mass point representation of the vehicle was implemented, which analyses only the longitudinal motion of the car. In general, there are two main approaches towards the longitudinal motion simulations of vehicles, kinematic and dynamic simulations. In case of a kinematic simulation, the actual state of the simulation components is calculated backward from a given driving condition, using the gear ratios and efficiencies to determine the required input values of the propulsion unit. In a dynamic simulation, where the calculations are forward-directed, a driver model calculates the torque demand of the vehicle from the current vehicle speed and provides the corresponding input to the propulsion unit. In the following blocks, the states of the drivetrain components are calculated, resulting in the vehicle's actual speed, which is connected to the driver model to close the simulation loop. Using MATLAB Simulink in the project, a dynamic simulation was implemented because this kind of model ensures a more realistic and accurate simulation of the vehicle's drivetrain than the kinematic approach (Winke and Bargende 2013). The simulation model is divided into three main blocks (see in Figure 5):

- Driver module
- Drivetrain module
- Vehicle module

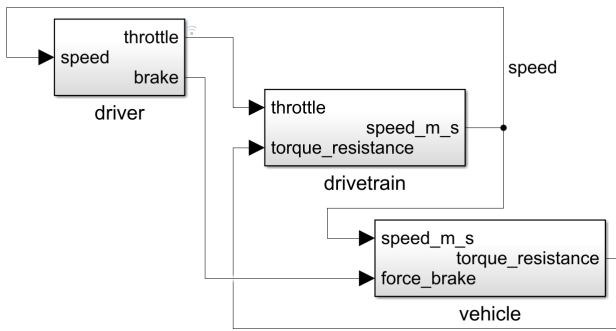


Fig. 5. The Basic Structure of the Dynamic Simulation Model

The *driver module* controls the electric motor's throttle based on the New European Driving Cycle (NEDC). If the current speed of the vehicle is lower than the desired, the module raises the throttle signal, if the opposite is true, it lowers the throttle signal, or even initiates braking.

The *drivetrain module* simulates the dynamics of the electric motor, the fixed-gear drive module, and the vehicle's wheel (see in Figure 6), using the throttle signal and the resistances acting on the vehicle as inputs to calculate the vehicle's speed. Each block calculates its inner state according to the input and output torques and angular accelerations, assuming that the system is totally stiff, and does not contain any non-linearity such as backlashes.

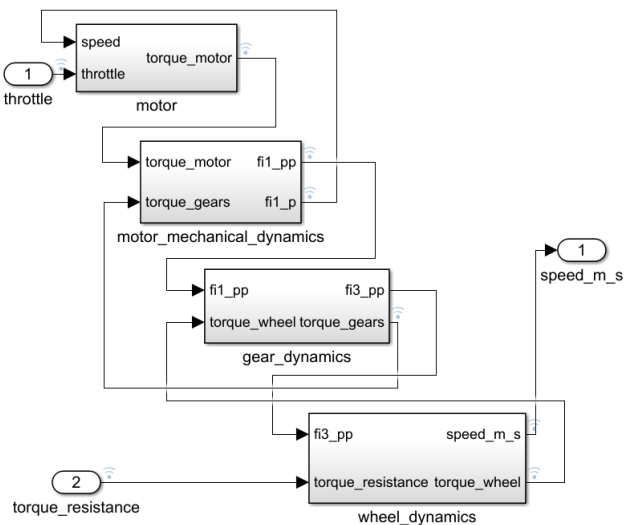


Fig. 6. Drivetrain Simulation Module's Block Diagram

Inside the *drivetrain module*, the *motor* block provides torque input to the *motor mechanical dynamics* block based on the current motor speed, the throttle signal (which can be between 0 and 1), and a look-up-table (LUT) which contains the motor's torque-speed characteristics. The LUT determines the maximal motor torque at the given speed; finally, this value is multiplied by the throttle signal value. This method enables to create the needed motor torque between zero and the nominal maximum torque of the motor for each motor speed. In the *motor mechanical dynamics* block, the angular acceleration of the motor is calculated as

shown in Equation (1):

$$\ddot{\varphi}_m(t) = \frac{\tau_{motor}(t) - \tau_{gears}(t)}{J_{motor}} \quad (1)$$

where,

- $\varphi_m$  is the motor's angular position,
- $\tau_{motor}$  is the motor torque,
- $\tau_{gears}$  is the input torque for the gearbox,
- $J_{motor}$  is the moment of inertia of the motor's rotating parts

Integral of the angular acceleration over time gives the motor speed, which is connected to the *motor* block.

The *gear dynamics* block consists of the input-, intermediate- and differential shaft assemblies, which are connected together by gear pairs (see in Figure 7).

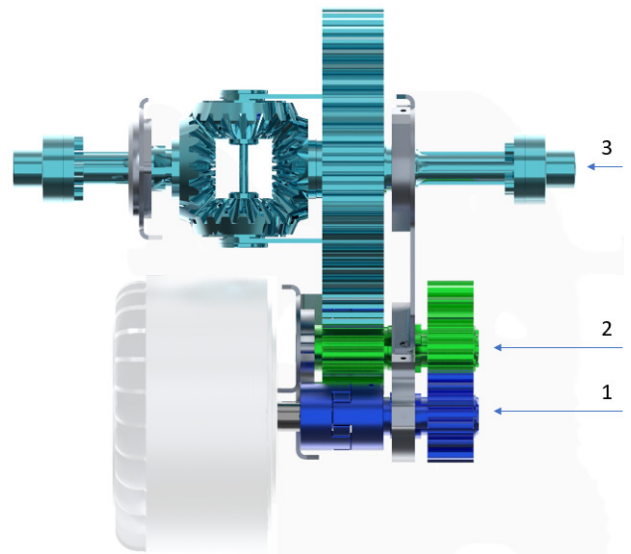


Fig. 7. 3D Model of the Fixed-gear Drive Module: Input- (1), Intermediate- (2), Differential Shaft (3) Assemblies

Since the motor is connected to the input shaft assembly, the angular acceleration of these two is equal. To ensure this the input torque for the gearbox ( $\tau_{gears}$ ) is calculated as shown in Equation (2):

$$\tau_{gears}(t) = \tau_{l1} + J_1 \cdot \ddot{\varphi}_m(t) + \frac{\tau_{l2} + J_2 \cdot \ddot{\varphi}_m(t) + \frac{\tau_{l3} + J_3 \cdot \ddot{\varphi}_m(t) + \tau_{wheels}(t)}{i_{2,3}}}{i_{1,2}} \quad (2)$$

where,

- $\tau_{wheels}$  is the input torque for the wheels
- $\tau_{l_j}$  is an estimated torque loss (e.g. from bearings) for the j-th shaft assembly,
- $J_j$  is the moment of inertia of the j-th shaft assembly,
- $i_{j,k}$  is the gear ratio of the gear pair between the j-th and the k-th shaft assemblies.

The *wheel dynamics* block represents the connection between the drivetrain's rotational- and the vehicle's longitudinal movement. The torque transferred

through the gearbox to the wheels is compensated by static friction on the ground; thus, the vehicle's center of gravity starts to accelerate in a longitudinal direction. Assuming that the contact points of the wheels are not sliding on the ground, the relation between the longitudinal and rotational measures can be expressed as shown in Equation (3) - (5):

$$a(t) = \ddot{\varphi}_{wheels}(t) \cdot r \quad (3)$$

$$m = \frac{J_{red}}{r^2} \quad (4)$$

$$F_P(t) = \frac{\tau_{red}(t)}{r} \quad (5)$$

where,

- $a$  is the vehicle's longitudinal acceleration,
- $\varphi_{wheels}$  is the wheels' angular position,
- $r$  is the wheel radius,
- $m$  is the vehicle's mass,
- $J_{red}$  is the reduced moment of inertia of the vehicle's mass,
- $F_P$  is the pulling force,
- $\tau_{red}$  is the torque that effectively accelerates the vehicle in longitudinal direction.

When the vehicle is moving straight, the wheels are rotating together at the same speed as the differential shaft assembly. Their angular acceleration can be calculated from the motor's angular acceleration (see in Equation (6)). Thus the torque acting on the wheels ( $\tau_{wheels}$ ) is responsible for the wheels' angular acceleration, the vehicle's longitudinal acceleration, and the compensation of resistance torques (see in Equation (7)).

$$\ddot{\varphi}_{wheels}(t) = \frac{\ddot{\varphi}_m(t)}{i_{1,2} \cdot i_{2,3}} \quad (6)$$

$$\tau_{wheels}(t) = (J_{wheels} + J_{red}) \cdot \ddot{\varphi}_{wheels}(t) + \tau_{res} \quad (7)$$

where,

- $J_{wheels}$  is the moment of inertia of the wheels,
- $\tau_{res}$  is the resultant torque from resistance forces acting on the vehicle.

Integral of Equation (3) over time gives the vehicle's longitudinal speed, which serves as an input in the *vehicle module* to calculate resistances, and also in the *driver module* to determine the throttle and brake signals.

The separation of the blocks allows us to change the current model into more detailed versions, add more gear stages, and to read out the inner states of each component throughout the simulation.

In the *vehicle module*, the resistance forces acting on the vehicle are simulated (see in Figure 8).

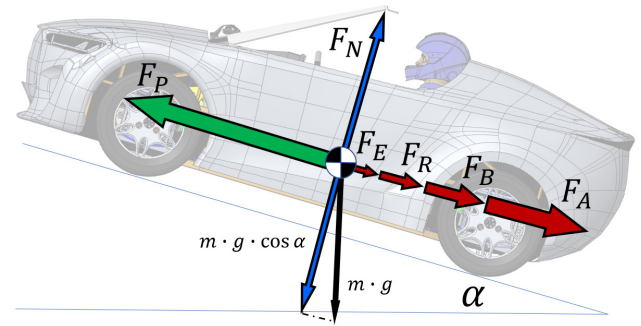


Fig. 8. Forces Acting on the Vehicle, at a Constant Velocity Equilibrium State

- $F_P$  : pulling force from  $\tau_{red}$ ,
- $F_A$  : force from air resistance,
- $F_R$  : force from rolling resistance,
- $F_E$  : force from elevation on the slope,
- $F_B$  : force from braking,
- $F_N$  : normal force,
- $g$  : gravitational acceleration,
- $\alpha$  : angle of slope.

For every equilibrium speed, the pulling force is equal to the summation of the resistance and elevation forces. If the pulling force is greater than the resistance forces, the vehicle is accelerating; else, it is decelerating.

Using this model a wide variety of valuable information can be extracted from the simulation results, such as the required motor torque to meet the prescribed speed profile (see in Figure 9) or the energy consumption during the examined time. This information can serve as input to other applications, which can check how good the final concept is. Modification of the parameters of mechanical parts is essential in the design process. Changes could be propagated through the simulation, and based on the results, optimal values could be calculated for the initial parameters. If we manipulate data in 3D models, a managing algorithm should keep track of the evaluated (or shared) parameters after the models are updated, to keep the Digital Twin consistent across CAD systems, simulation programs, or any other platforms.

### Managing algorithm

Usually, data exchange between different applications from different fields (e.g. 3D modeling, simulation, optimization) is not provided. It requires some programming skills to extract data from one in a form that is useful to the other. Each of these software products rely heavily on graphic interfaces, however, if the models are appropriately set up, both of them can be managed from a third program. This managing program could enable the parameter optimization of the product; in our case, the drivetrain parameters can be optimized to minimize energy consumption.

If access to the model and the simulation results is provided, an optimizer algorithm could tune parameters in order to reach optimal properties of the modeled

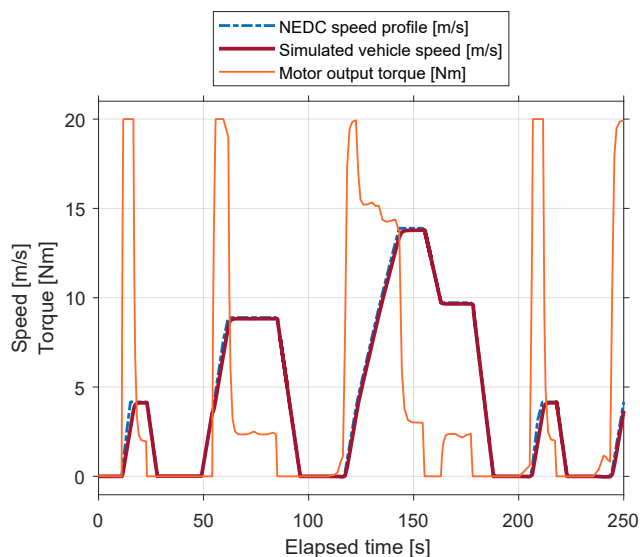


Fig. 9. Visualization of Simulation Results: Required Motor Output Torque to ensure the Prescribed Speed Profile for the Vehicle

system, however such an optimizer algorithm needs to be implemented in this project in the future. In the project, a simple Python script is used to ensure communication between Creo and Simulink.

## CONCLUSIONS

The concept of the Digital Twin and Generative Design was introduced with its benefits and possibilities. Then small electric vehicles' advantages and disadvantages were reviewed. In the example project, the 3D CAD model, the simulation model, and the drive module's mathematical model were detailed.

## FURTHER DEVELOPMENT

The detailed 3D model of the drive module and a simple simulation model of the powertrain is set up and ready to be utilized for measurements and tests. The model parameters and results could be accessed from a Python script, through which other different programs could be used to analyze the product from different aspects. As this managing algorithm could handle parameters consistently, it is capable of implementing an optimization algorithm to generate various optimal solutions for a defined problem. In the future, implementing a sensor network on the drive module to measure the actual torques and velocities, to identify the estimated loss parameters accurately, to test different scenarios and validate the simulation model is necessary. Based on these corrected parameters, the drivetrain module could be tested in dangerous situations virtually without damaging or breaking the physical twin. With a proper optimization algorithm (e.g. genetic algorithm), geometrical parameters could be optimized. Later, the model completed with the vehicle's suspension system could be used in a 3D physical engine for further optimization and testing.

## ACKNOWLEDGEMENT

The research reported in this paper was supported by the Higher Education Excellence Program of the Ministry of Human Capacities in the frame of Artificial intelligence research area of Budapest University of Technology and Economics (BME FIKP-MI), and by the National Research, Development and Innovation Fund (TUDFO/51757/2019-ITM, Thematic Excellence Program).

## REFERENCES

- Goossens, P. (2017). Industry 4.0 and the power of the digital twin. Maplesoft Engineering Solutions.
- Grieves, M., & Vickers, J. (2017). Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. [https://doi.org/10.1007/978-3-319-38756-7\\_4](https://doi.org/10.1007/978-3-319-38756-7_4)
- Khan, S., & Awan, M. (2018). A generative design technique for exploring shape variations. *Advanced Engineering Informatics*, 38, 712–724. <https://doi.org/10.1016/j.aei.2018.10.005>
- Park, G., Lee, S., Jin, S., & Kwak, S. (2014). Integrated modeling and analysis of dynamics for electric vehicle powertrains. *Expert Systems with Applications: An International Journal*, 41, 2595–2607. <https://doi.org/10.1016/j.eswa.2013.10.007>
- Rodič, B. (2017). Industry 4.0 and the new simulation modelling paradigm. *Organizacija*, 50, 193–207. <https://doi.org/10.1515/orga-2017-0017>
- Un-Noor, F., Sanjeevikumar, P., Mihet-Popa, L., Molah, M., & Hossain, E. (2017). A comprehensive study of key electric vehicle (ev) components, technologies, challenges, impacts, and future direction of development. *Energies*, 10. <https://doi.org/10.3390/en10081217>
- Winke, F., & Bargende, M. (2013). Dynamic Simulation of Urban Hybrid Electric Vehicles. *MTZ worldwide*, 74, 56–63. <https://doi.org/10.1007/s38313-013-0089-1>



**TAMÁS DÓKA** is a PhD student at the Department of Machine and Product Design of Budapest University of Technology and Economics (BME), Hungary. He had received masters degree in Mechatronic Engineering in 2019. He is actively taking part in an Electric Vehicle development university project called BME Fuse, focusing on concept creating, 3D modeling, simulation and optimization. Email address: doka.tamas@gt3.bme.hu



**PÉTER HORÁK, PhD** is an associate professor at the Department of Machine and Product Design of Budapest University of Technology and Economics (BME), Hungary. He had received PhD degree in Mechanical Engineering in 2004. His research fields are tribology, kinematics and geometry of gears, design theory and methodology. Email address: horak.peter@gt3.bme.hu

# SIMULATION-BASED EVALUATION OF RESERVATION MECHANISMS FOR THE TIME WINDOW ROUTING METHOD

Thomas Lienert

Florian Wenzler

Johannes Fottner

Chair of Materials Handling, Material Flow, Logistics

Department of Mechanical Engineering

Technical University of Munich

Boltzmannstrasse 15, 85748 Garching, Germany

Email: thomas.lienert@tum.de, florian.wenzler@tum.de, j.fottner@tum.de

## KEYWORDS

Automated Warehouses, Mobile Robots, Time Window Routing Method, Discrete Event Simulation

## ABSTRACT

Automated warehouses operated by a fleet of robots offer great flexibility, since fleet size can be adjusted easily to throughput requirements. Furthermore, they provide higher redundancy compared to common solutions for automated storage and retrieval systems.

On the other hand, these systems require more complex control strategies to run robustly and efficiently. Special routing and deadlock handling strategies are necessary to avoid blocking and collisions among the robots.

In this contribution, we focus on the time window routing method, an approach for avoiding deadlocks by reserving routes in advance. We present and discuss different reservation mechanisms that are evaluated by the means of simulation.

## INTRODUCTION

Automated warehouses that are run with a fleet of mobile robots for the part-to-picker order picking have become the subject of intensive research since different decision problems need to be resolved during operation of these systems.

Basically, these systems consist of a rack system containing storage items and a fleet of robots moving within the storage area. The robots use a rectangular grid of paths to fulfil storage and retrieval requests. We refer to these systems as mobile-robot-based warehouses. There are several types of mobile-robot-based warehouses that differ in regard to their storage systems.

Robotic mobile fulfilment systems (RMFS) consist of a single storage tier, where items are stored on shelves on the ground. Robots travel underneath these shelves, lift them and bring them to the picking zone that is located somewhere near the storage area (Azadeh et al. 2018a).

In contrast, shuttle-systems consist of several tiers that are connected by lifts. These lifts link the storage system to the picking zone and enable vertical movements of the robots that are denominated as shuttles in this context (Tappia et al. 2018).

In another type, robots move horizontally as well as vertically within an aisle placed between two single-deep storage racks. Picking zones are located at one or at both ends of each aisle (Azadeh et al. 2018b).

Although these types differ slightly, they provide the same benefits relative to common stacker-crane-based storage systems. They are easily scalable. The whole system can theoretically be run with a single robot and if a higher throughput is needed, more and more robots can be added. The layout can be changed flexibly and the system can be enlarged easily. Furthermore, a required sequence can be established within the storage system (Lienert and Fottner 2018) and high redundancy can be provided as long as suitable failure-handling strategies are used (Lienert et al. 2019). In the literature, mobile-robot-based warehouses are widely discussed. Among others, storage assignment (Boysen et al. 2019), order batching (Boysen et al. 2017), dispatching (Yuan and Gong 2017), battery charging and swapping (Zou et al. 2017) as well as dwelling strategies for idle robots (Roy et al. 2016) are addressed.

In this contribution, we focus on the routing and deadlock-handling, more precisely on the time window routing method. This approach enables conflict-free routing of robots by reserving the path to be travelled in advance. Acceleration and deceleration processes are usually neglected when the time window routing method is applied. We present different reservation mechanisms that include acceleration and deceleration processes and fit different requirements regarding communication between the robots and the material flow control. The remainder of this paper is organized as follows. In the next section, we briefly introduce the time window routing method. We subsequently describe and discuss three different reservation mechanisms that are compared with a simulation and taking into account an RMFS before we conclude our work.

## TIME WINDOW ROUTING METHOD

In mobile-robot-based warehouses, robots move using the layout given by storage-aisles and cross-aisles. Since several robots are operating in the system at the same time, traffic must somehow be controlled and deadlocks or even worse, collisions must be avoided. In general a deadlock describes a situation where one or more



processes are blocked forever because the requests for resources by the processes can never be satisfied (Kim et al. 1997). In the context of routing robots, the processes correspond to the execution of the routes and the resources for the layout of segments along these routes. For example, a deadlock occurs if two robots driving in opposite directions meet each other within a storage aisle.

The approach of the time window routing method avoids deadlocks by reserving the path for a robot from its current situation to the destination in advance. On each layout segment that needs to be travelled along this path, a time window is blocked, during which the layout segment is claimed exclusively by a robot and during which the layout is not available for the movement of any other robot. Since time windows on neighbouring layout segments of a route overlap each other, robots can move safely through the layout.

To apply this method, the layout is represented by a graph. Each node corresponds to a layout segment, whereas the edges give information about predecessors and successors of the nodes. For each node, there is a time line with reserved and free time windows (figure 1).  $f_{i,l}$  denominates the  $l^{\text{th}}$  free time window on node  $i$ .

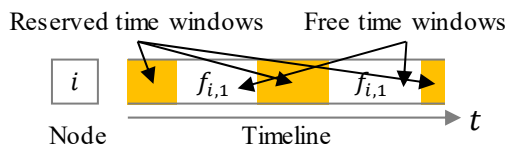


Figure 1: Reserved and free time windows on a node  $i$

In the case that a robot has to be routed from its current position to a given destination, the method searches for a route through the free time windows on the nodes using an A\*-algorithm. Once a conflict-free route is found, the corresponding time windows are reserved and the robot can start moving. From the point of time that the idea of this method was introduced first by Kim and Tanchocco (Kim and Tanchocco 1991), it has been applied in different contexts as the routing of automated guided vehicles in container terminals (Stenzel 2008) for the organization of taxi traffic at airports (Bussacker 2005) and for managing a fleet of robots in an RMFS (Hvězda et al. 2018). For more detailed insights into the modelling of the layout as a graph we refer to (Lienert and Fottner 2017).

The time window routing method can be applied in systems that use centralized material flow control as well as in systems that are operated by a multi agent system.

When it comes to the execution of a reserved route, deadlocks or collisions among the robots threaten to occur even though the routes are theoretically conflict-free. Robots might be delayed – due to several reasons – and not match their reserved time windows. Therefore it is essential that the node’s crossing order of the robots based on the conflict-free schedule is maintained (Maza and Castagna 2005). Hence, a robot is only allowed to travel the next node along its reserved route if it has reserved the next time window on that node.

Figure 2 shows an example that clarifies this approach. At timestamp  $T$  the reserved time window of robot  $r_2$  begins on node  $j$ . Robot  $r_1$  is delayed. According to the planning, it should reside on node  $i$ , but it has not yet passed node  $j$ ,  $k$  and  $l$ .

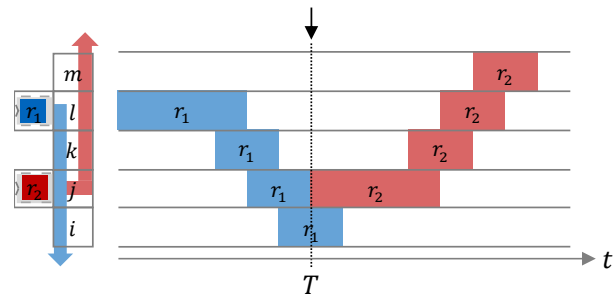


Figure 2: Robot  $r_1$  is delayed and a deadlock might occur

If robot  $r_2$  nevertheless continues with the execution of its route, both robots will face each other somewhere in between the nodes  $j$  and  $l$  and a deadlock will occur.

Note that robots must also not enter a node before their reserved time window has started. In the example in figure 3, robot  $r_2$  has just entered node  $j$  at timestamp  $T$  and before the corresponding time window has started. Routing robot  $r_1$  will lead to a feasible route traversing nodes  $l$ ,  $k$  and  $j$  before the reservation of robot  $r_2$  begins (as in figure 2). Once again, a deadlock is likely to happen.

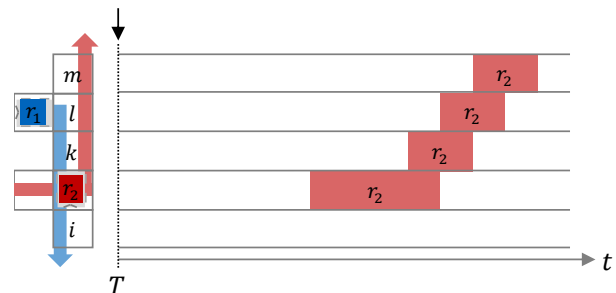


Figure 3: Robot  $r_2$  is early and a deadlock might occur

Maintaining the node’s crossing order is easily realizable if acceleration and deceleration processes are neglected since a robot can stop immediately in case it is not allowed to enter the next node. However, if acceleration and deceleration processes are taken into account, the robust execution of a route must be implemented with some sort of lookahead.

In a previous work, we modified the time window routing method to incorporate acceleration and deceleration processes. During the planning, so-called “segments” are created that describe movement of a robot over several nodes in a straight line. The computed route is executed segment by segment, respecting the node’s crossing order (Lienert et al. 2018a). Figure 4 shows the creation of a segment during the planning phase. Starting with a free time window on the node  $i$ , the algorithm checks whether the free time window on node  $j$  is reachable. In that case, the segment is – if possible – extended node by node.

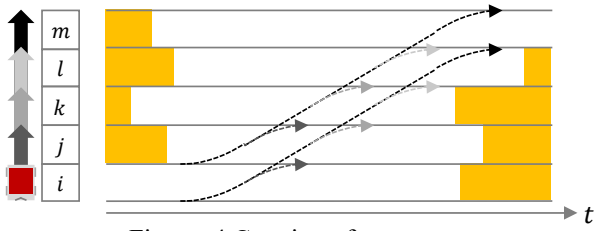


Figure 4: Creation of a segment.

To extend a segment, none of the time windows within the segment can violate any existing reservation. In the example in figure 4, all free time windows are reachable since no trajectory touches any existing reservation. Each trajectory is described by a pair of arrows that represent the movements of the front and rear of a robot. Free time windows that are reachable are candidates for the first node of a new segment in a later iteration. For a comprehensive description of the overall routing procedure, we refer to (Lienert et al. 2018a).

### RESERVATION MECHANISMS

In this section, we focus on the part of the algorithm that examines the reachability of free time windows from a free time window at a specific time stamp, taking into account three different reservation mechanisms. These mechanisms differ on the one hand in regard to the requirements of the communication between robots and centralized material flow control. On the other hand, the required length of the reserved time windows of a segment differ, which leads to varying resource utilizations.

#### Mechanism 1: Triangle

First, we assume that once a robot starts with the execution of a segment, it has to be ensured that the robot can finish that segment without any interference due to other robots that are late. In settings where communication between robots and central material flow control cannot be guaranteed at any given time in real-time, this procedure is necessary to avoid collisions and deadlocks.

The reservations of a segment are preponed so they start with the departure of the robot from the first node of the segment. Time windows are deleted as soon as a robot has left a node completely. As a result, reservations of a segment form a triangle (see figure 5).

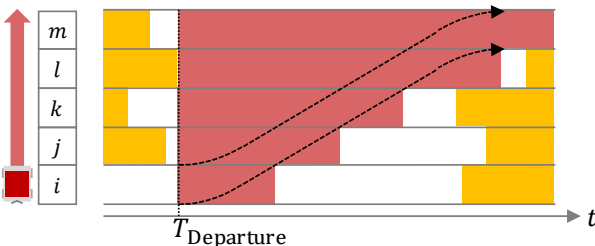


Figure 5: Reservations of a segment display the form of a triangle

We assume free time window  $f_{i,l}$  to be the first time window of a segment and free time window  $f_{k,p}$  the free time window whose reachability is analyzed (figure 6).

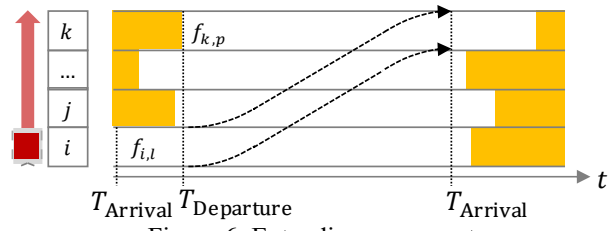


Figure 6: Extending a segment

The flowchart in figure 7 shows the procedure of checking the reachability of a free time window (FTW) in detail.

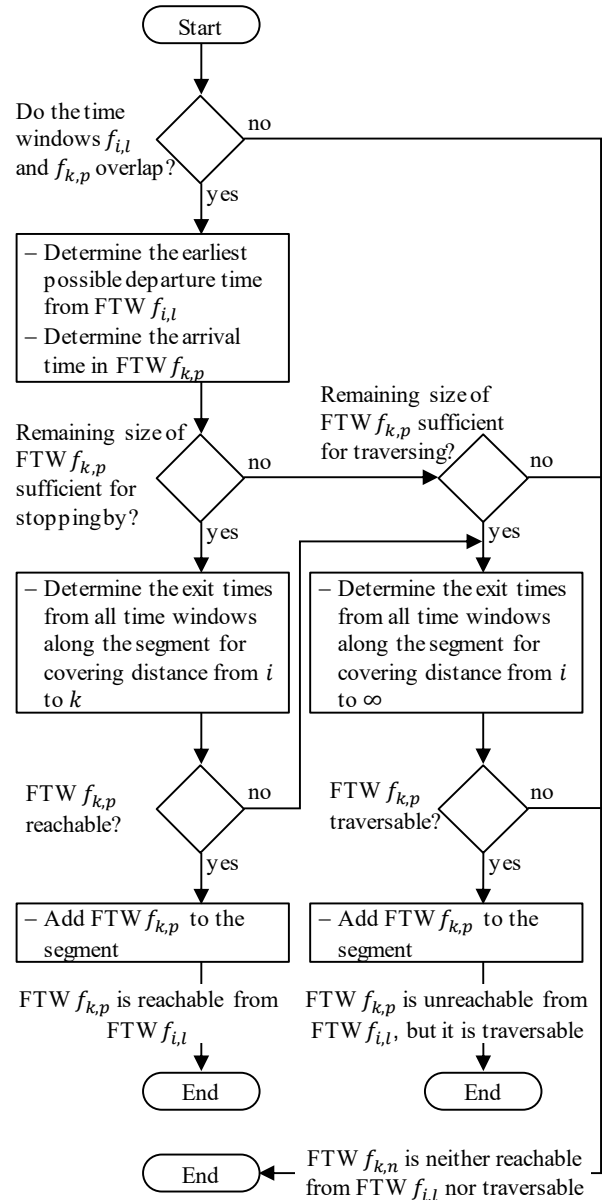


Figure 7: Reachability check of free time window  $f_{k,p}$

First of all, it is mandatory that both of the free time windows  $f_{i,l}$  and  $f_{k,p}$  overlap each other. In that case, the earliest departure time from free time window  $f_{i,l}$  can be determined. The departure cannot take place before the robot arrives on node  $i$  and not before any of the free time windows of the segment including the free

time window  $f_{k,p}$  start. Next, the arrival time in free time window  $f_{k,p}$  can be calculated taking into account the distance as well as the robot's parameters.

The remaining size of free time window  $f_{k,p}$  after the arrival must be sufficient to leave the node completely before the next reservation starts. In that case, all of the required reservations on all nodes of the segment can be determined. If no existing reservation on any node of the segment is violated, the free time window  $f_{k,p}$  is reachable and the segment can be extended. In a later iteration, free time window  $f_{k,p}$  serves as a starting time window of another segment.

If the remaining size of free time window  $f_{k,p}$  is not sufficient for stopping by, node  $k$  might be traversed by the robot before the next reservation starts. In that case, all required time windows on all nodes of the segment can once again be determined assuming a movement of infinite length. If no existing reservation on any node of the segment is violated, free time window  $f_{k,p}$  is traversable and the segment can be extended (as in figure 8).

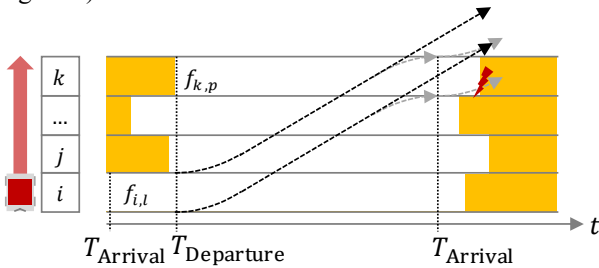


Figure 8: Remaining size of free time window  $f_{k,p}$  does not allow an intermediate stop on node  $k$ , but traversing the node.

In case the remaining size of the free time window is sufficient for stopping by but a previous reservation is violated, traversing the node without violating any reservation might once again be possible (as in figure 9).

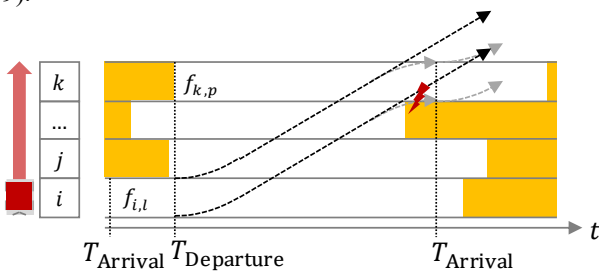


Figure 9: An existing reservation does not allow an intermediate stop on node  $k$ , but does allow traversing the node.

Note that in both cases (figure 8 and figure 9), free time window  $f_{k,p}$  is not a potential candidate for the first time window of another segment.

During execution of a reserved route, a robot is only allowed to start a segment if the robot is not early and if the robot is to travel all nodes of the segment next, taking into account the node's crossing order.

## Mechanism 2: Rectangle

The second reservation mechanism is even more restrictive. Time windows are deleted only after a segment is finished and the robot comes to a standstill. As a result, reservations of a segment form a rectangle (see figure 10).

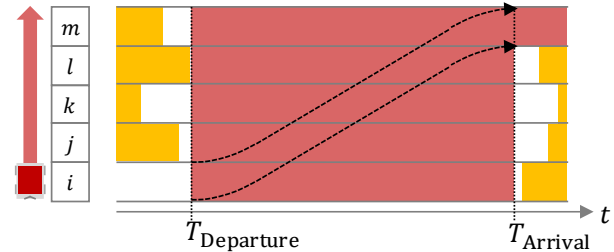


Figure 10: Reservations of a segment display the form of a rectangle

This approach is suitable for applications that do not allow the deletion of time windows in real-time. Once a robot starts with the execution of a segment, no information of the robot's position is available until the robot finishes the segment.

The procedure to check reachability corresponds to the flow chart in figure 7 apart from determining the necessary reservations. These once again begin with the departure, but end with the arrival on the last node of the segment.

## Mechanism 3: Stairs

The third reservation mechanism enables a more efficient use of resources but requires communication between robots and centralized material flow control more often. Time windows do not begin at the departure, but there is a buffer before each reservation that enables a safety deceleration in case a robot with a previous reservation is delayed. As a result, reservations of a segment form a stairs pattern (see figure 11).

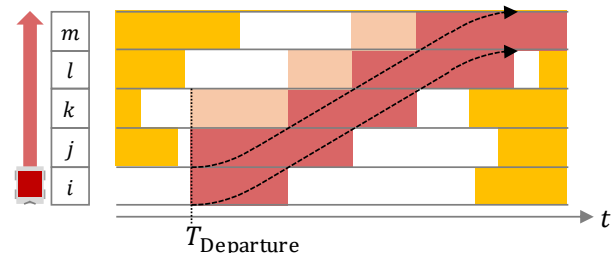


Figure 11: Reservations of a segment display the form of a stairs pattern.

To analyze reachability, the procedure differs slightly and is shown in the flow chart in figure 14. Note that it is not necessary that both free time windows  $f_{i,j}$  and  $f_{k,p}$  overlap each other.

First of all, the start of the reservation of the free time window  $f_{k,p}$  has to be determined. Starting from the penultimate node of the segment on which the robot has to stop in case a preceding robot is delayed, the

distances between the nodes are added up until the deceleration distance is reached. Let the node  $a$  be the node in which this summed distance reaches or exceeds the deceleration distance. The reservation of the free time window  $f_{k,p}$  must begin as soon as the stopping position of the node  $a$  has been passed (figure 12).

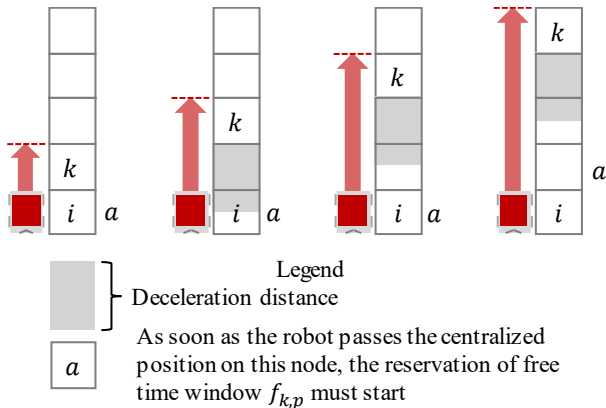


Figure 12: Determination of the start of the reservation of the free time-window  $f_{k,p}$

If the calculated start does not fall in the free time-window, it is checked whether it is possible to postpone the departure, so that the start coincides with the start of the free time window. Otherwise the free time window  $f_{k,p}$  is not reachable.

The arrival time can be determined next. The remainder of the procedure corresponds to the one described in the flowchart in figure 7 apart from the determination of the necessary reservations. For each free time window in the segment, the start of the reservations needs to be recalculated (as described above) since the departure from the first free time window  $f_{i,j}$  of the segment might have been shifted. Reservations end as soon as a robot has completely left a node.

In case a robot that reserved a preceding time window on any node of a segment is delayed, the additional buffer allows the robot to decelerate and to stop safely. In the example in figure 13, robot  $r_1$  is delayed at timestamp  $T$  and its reserved time window on node  $m$  is not yet deleted, robot  $r_2$  is not allowed to enter node  $m$  and starts decelerating, coming to a standstill on node  $l$ .

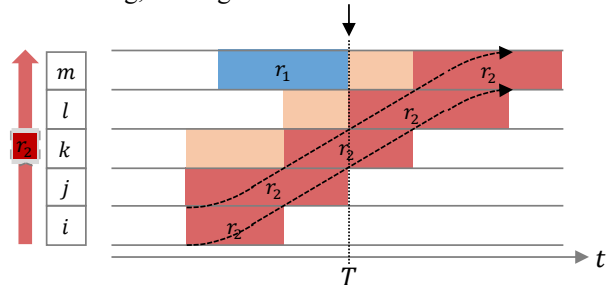


Figure 13: Robot  $r_1$  being late leads to an unplanned intermediate stop on node  $l$  for robot  $r_2$ .

Note that if acceleration and deceleration processes are neglected, additional buffers disappear.

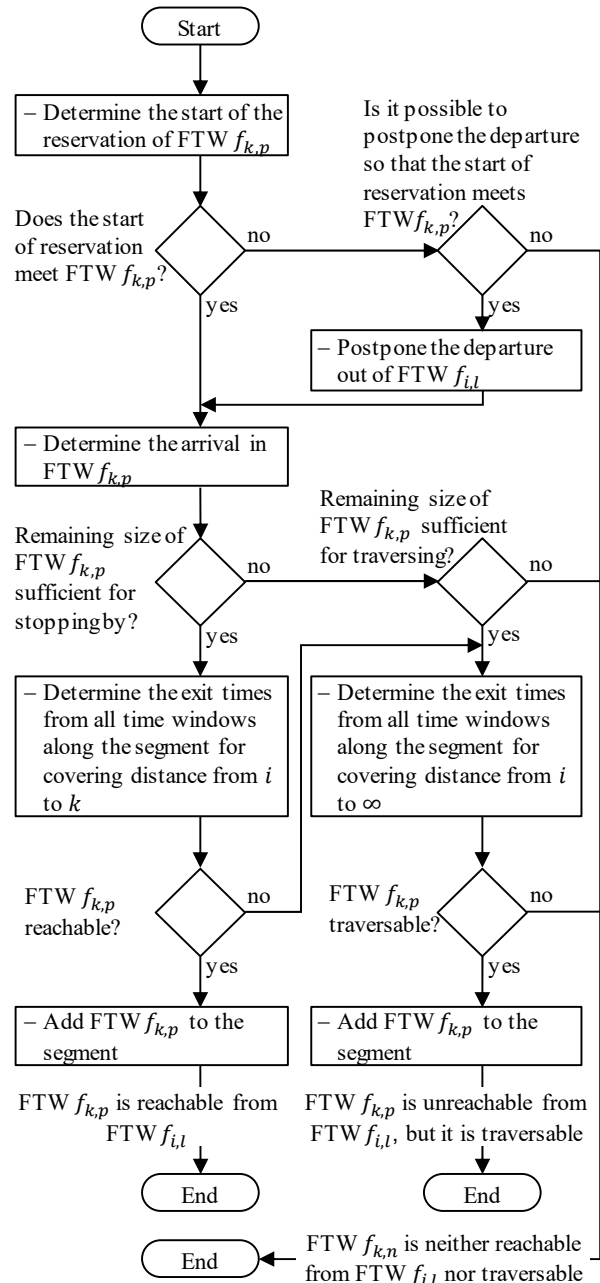


Figure 14: Reachability check of free time window  $f_{k,p}$

Theoretically there is another reservation mechanism where reservations start with an additional buffer but end with the arrival on the last node of the segment. However, we assume that if real-time communication is enabled (which is necessary for the stairs mechanism), time windows can be deleted immediately after a node has been left.

## SIMULATION STUDY

In this section, we compare the previously described reservation mechanisms by performing a simulation study, considering an RMFS.

There is a trade-off regarding the segment length using the triangle and rectangle reservation mechanisms. Short segments entail frequent stops and accelerations of the robots. Longer segments enable robots to achieve

maximum speed and reach their destinations with fewer intermediate stops. But in this case, the reserved time windows on the nodes are larger and the nodes are blocked longer for other vehicles. As a consequence, the optimal maximal segment length has to be determined for each number of robots. This is done first before the mechanisms are compared to each other.

### Considered System

We apply the strategies to a fleet of robots moving within an RMFS with 336 storage locations that are arranged in seven double rows divided by storage aisles. There are two cross-aisles located at one third and at two thirds of the aisle length. All aisles can be used for bi-directional traffic. There are four picking areas with five picking places, with each arranged in front of the storage system. In front of these places, there are two unidirectional cross-aisles. A replenishment area, where empty racks are refilled, is located on the opposite side of the storage area. Robots are dedicated to a picking zone and perform three different cycles to maintain the material flow between storage locations, picking area and replenishment area. For a more detailed description of the system, we refer to (Lienert et al. 2018b). We implemented the RMFS using the Tecnomatix Plant Simulation discrete event simulation environment.

### Parameter Settings

We vary the number of robots, starting with four robots (one for each picking zone) and increasing this up to 60 robots working in the system in steps of four, and repeat the experiments for each reservation mechanism. With both the triangle and rectangle reservation mechanisms, we limit the segment length to a certain number of nodes. We start with a maximum segment length of only one node, meaning that robots only move node by node, stopping at every single node along their routes. We vary the maximum segment length increasing it up to 30 nodes. All of the remaining parameters, such as the robot's acceleration and maximum speed, remain the same. Simulation time is set to 24 hours. No warm-up time is taken into account, as the goal of the simulation study is to compare the reservation mechanisms. We conduct five replications for each parameter setting.

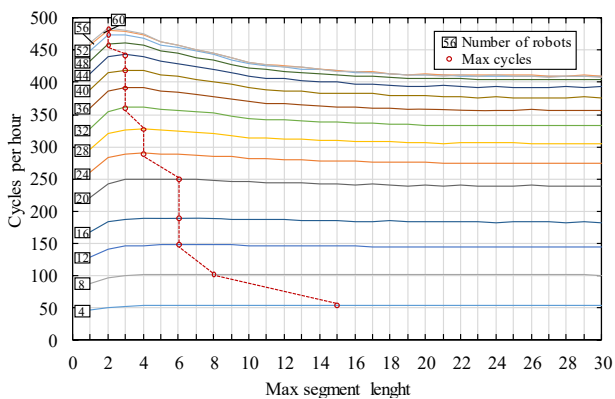


Figure 15: Throughput using the triangle reservation mechanism

### Results

First we analyse the trade-off between shorter and longer segments. Figure 15 shows the throughput reached with the corresponding parameter setting using the triangle reservation mechanism.

As can be seen, the more robots that are operating in the system, the shorter the optimal segment length becomes. Figure 16 shows the throughput reached using the rectangle reservation mechanism.

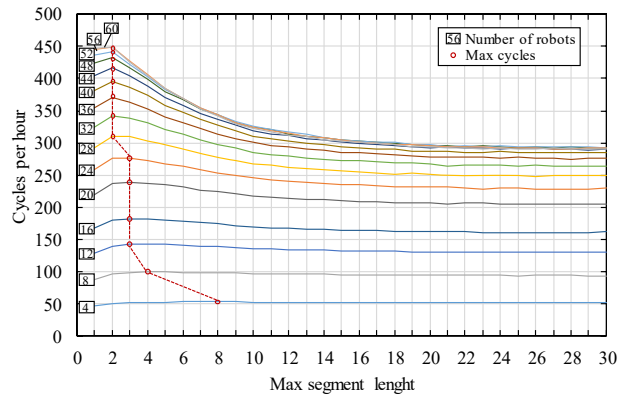


Figure 16: Throughput using the rectangle reservation mechanism

A similar behavior can be observed. However, the optimal maximum segment length is even shorter. That is expectable, since nodes are reserved for a longer time periods than using the triangle mechanisms.

Finally we compare these maximums to the stairs reservation mechanisms. Figure 17 shows the throughput reached.

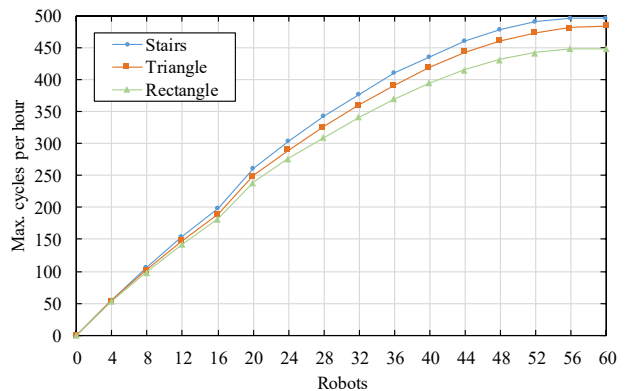


Figure 17: Throughput for varying number of robots

All curves look similar with a linear increase and a small knee between 16 and 20 vehicles before reaching saturation with 56 robots. Since each robot is assigned to a certain picking zone, the number of robots per picking zone with 20 robots equals the number of picking places, and a different strategy is used for the supply of the picking zone (see Lienert et al. 2018b).

For a small number of robots, the throughput of the different mechanisms is equal. In fact, using a single robot in the whole system yields the same throughput, no matter which reservation mechanism is used.

However, the more robots that are operating in the system, the more evident the difference becomes. As expected, the stairs mechanism reaches the highest throughput followed by the triangle one. Reaching a throughput of 400 cycles per hour using the stairs mechanism requires 36 robots, 40 robots using the triangle mechanism and as many as 44 robots using the rectangle mechanism.

Note that saturation is reached at different levels. Using the stairs reservation mechanism yields the highest throughput, whereas the triangle mechanism reaches 97.6% and the rectangle mechanism 90.4% of that throughput.

## CONCLUSION

In this contribution, we considered mobile-robot-based warehouses. We presented three different reservation mechanisms for the time window routing method. These mechanisms require different levels of communication and differ regarding the resource utilization while executing a calculated route. We conducted a simulation study to compare the performance of these mechanisms considering an RMFS. As expected, the stairs mechanism enables the highest throughput followed by the rectangle mechanism.

In our consideration, robots accelerate, move with constant maximum speed or decelerate. For future work, we suggest enabling robots to move with a constant, but reduced speed to avoid intermediate stops. This behavior must be modelled and taken into account for the analysis of reachability of free time windows.

## REFERENCES

- Azadeh, K.; de Koster, R. and Roy, D., 2018 (a), "Robotized and Automated Warehouse Systems: Review and Recent Developments." *Transportation Science* 53 No.4, 917–945.
- Azadeh, K.; Roy, D.; De Koster, R., 2018 (b): "Design, Modeling, and Analysis of Vertical Robotic Storage and Retrieval Systems." *Transportation Science* 53 No.5, 1–22.
- Boysen, N.; Briskorn, D. and Emde, S., 2017, "Parts-to-picker based order processing in a rack-moving mobile robots environment." *European Journal of Operational Research* 262, No.2, 550-562.
- Boysen, N.; de Koster, R. and Weidinger, F., 2019, "Warehousing in the e-commerce era: A survey." *European Journal of Operational Research* 227 No.2, 396–411.
- Busacker, T. 2005. *Steigerung der Flughafen-Kapazität durch Modellierung und Optimierung von Flughafen-Boden-Rollverkehr – Ein Beitrag zu einem künftigen Rollführungssystem*. Dissertation. Technische Universität Berlin.
- Havězda, J.; Rybecký, T.; Kulich, M. and Přeučil, L., 2018, "Context-Aware Route Planning for Automated Warehouses." In *Proceedings of the 21st International Conference on Intelligent Transportation Systems*, 2955-2960.
- Kim C. W. and Tanchoco J. M. A., 1991, "Conflict-free shortest-time bi-directional AGV routing." *International Journal of Production Research* 29, No.12, 2377-2391.
- Kim C. W., Tanchoco J. M. A. and Koo P., 1997 "Deadlock Prevention in Manufacturing Systems with AGV Systems: Banker's Algorithm Approach." *Journal of Manufacturing Science and Engineering* 119, No.4, 849-854.
- Lienert, T. and Fottner, J., 2017, "Development of a generic simulation method for the time window routing of automated guided vehicles." *Logistics Journal: Proceedings*, Vol. 2017.
- Lienert, T.; Wenzler, F. and Fottner, J., 2018 (a), "Robust integration of acceleration and deceleration processes into the time window routing method." In *Proceedings of the 9th International Scientific Symposium on Logistics*, 66-86.
- Lienert, T.; Staab, T.; Ludwig, C. and Fottner, J., 2018 (b), "Simulation-based Performance Analysis in Robotic Mobile Fulfillment Systems." In *Proceedings of the 8th International Conference on Simulation and Modeling Methodologies, Technologies and Applications*, 383-390.
- Lienert, T. and Fottner, J., 2018, "Routing-based Sequencing Applied to Shuttle Systems." In *Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2949-2954.
- Lienert, T., Stigler, L. and Fottner, J., 2019, "Failure-handling strategies for mobile robots in automated warehouses." In *Proceedings of the 33rd European Conference on Modelling and Simulation*, 199-205.
- Maza, S. and Castagna, P., 2005, "A performance-based structural policy for conflict-free routing of bi-directional automated guided vehicles." *Computers in Industry* 56, No.7, 719-733.
- Roy, D.; Krishnamurthy, A.; Heragu, S. and Malmborg, C., 2015, "Queuing models to analyze dwell-point and cross-aisle location in autonomous vehicle-based warehouse systems", *European Journal of Operational Research* 242, No. 1, 72-87.
- Stenzel, B. 2008. *Online Disjoint Vehicle Routing with Application to AGV Routing*. Dissertation. Technische Universität Berlin.
- Tappia, E.; Roy, D. de Koster, R. and Melacini, M., 2018, "Modeling, Analysis, and Design Insights for Shuttle-Based Compact Storage Systems", *Transportation Science* 51, No.1.
- Yuan, Z.; Gong, Y.Y., 2017, "Bot-In-Time Delivery for Robotic Mobile Fulfillment Systems." *IEEE Transactions on Engineering and Management* 64, No.1, 83-93.
- Zou, B.; Xu, X.; Gong, Y.Y., de Koster, R., 2017, "Evaluating battery charging and swapping strategies in a robotic mobile fulfillment system." *European Journal of Operational Research* 267, No. 2, 733-753.

**THOMAS LIENERT** has been working as a research assistant at the Chair of Materials Handling, Material Flow and Logistics, Technical University of Munich, since 2014. His research deals with the simulation of mobile-robot-based warehouses. His email address is: thomas.lienert@tum.de.

**FLORIAN WENZLER** has been working as a research assistant at the Chair of Materials Handling, Material Flow and Logistics, Technical University of Munich, since 2013. His research deals with the optimization of project plans in construction. His email address is: florian.wenzler@tum.de.

**JOHANNES FOTTNER** is professor and head of the Chair of Materials Handling, Material flow, Logistics at the Technical University of Munich. His email address is: j.fottner@tum.de.

# MATHEMATICAL SIMULATION OF ADJACENT-COUPLING AMMONIA ABSORPTIVE REACTOR.

Wenchan Qi  
René Bañares-Alcántara  
Department of Engineering Science  
University of Oxford  
Parks Road, Oxford OX1 3PJ, UK  
E-mail: wenchan.qi@eng.ox.ac.uk

## KEYWORDS

Ammonia synthesis, mathematical model, absorption, Backflow Cell Model, absorptive reactor, adjacent-coupling

## ABSTRACT

The development of an efficient process for ammonia synthesis is a goal that has been long sought after; therefore, the application of an absorptive reactor for ammonia synthesis is important since it allows the reaction to occur under milder conditions. In the adjacent-coupling absorptive reactor, absorbent particles are positioned downstream the fixed ammonia synthesis catalyst bed. This kind of absorptive reactor leads to the enhanced conversion of ammonia synthesis under milder conditions, compared to the equivalent reactor used without absorbent. Here, we present the transient backflow cell model (BCM) to explain and analyse the phenomenon of absorption-enhanced reaction. The transient BCM, based on the first principle of mass balances, is developed to simulate that the backflow existing through the absorptive reactor. As a reference, the transient cell model (CM) is also implemented to simulate the absorptive reactor when assuming no existing backflow existing. These two models demonstrated that backflow through the absorptive reactor promotes the ammonia reaction conversion via two mechanisms: longer residence time for reaction and faster reaction rate due to the absorption of ammonia absorbed.

## 1. INTRODUCTION

Ammonia synthesis through the Haber-Bosch process is widely recognised as one of the most significant industrial applications (Bruce and Faunce 2015). Gas reactants hydrogen and nitrogen are converted to ammonia, operating in a catalytic packed bed reactor under high pressure (150-250 bar) and high temperature (400-500°C) (Appl 1999). Due to the problems of low single-pass conversion and energy-intensive operating conditions, the development of an ammonia synthesis reactor to enhance single-pass product yield and to adapt technology to and a less energy-intensive operation has attracted much attention.

A small number of research studies have investigated using the absorptive reactor to improve ammonia synthesis (Huberty et al. 2012; Wagner et al. 2017; Smith et al. 2019). The ammonia synthesis was enhanced by magnesium chloride absorption (Huberty et al. 2012), as the composite absorbents based on chlorides of alkaline-earth metals have a high ammonia absorption capacity (Zhu et al. 2009). Column absorption for reproducible cyclic separation in small scale ammonia synthesis was introduced in (Wagner et al. 2017). C. Smith et al. even ran the reaction and absorption at one temperature in the same piece of process equipment (Smith et al. 2019). However, no research focuses on the case when the catalyst bed and absorbent beds are sequentially packed in one column. In this paper, we focused on this adjacent-coupling configuration of the absorptive reactor for ammonia synthesis. As an increase in conversion was measured experimentally, it can be concluded that the flow regime is not an ideal plug flow, but backflow exists through the reactor. In this paper, the backflow is proposed to explain the absorption-enhanced conversion of ammonia synthesis. The reason for the improvement of the conversion will be identified in this paper. The transient backflow cell model (BCM) for the absorptive reactor was created to specify or verify the extent of absorption effect and determine the most efficient absorbent amount. The backflow cell model has actually been adopted in many types of research before (McSwain and Durbin 1966; Sinkule et al. 1976). The backflow cell model is of convenient because the mathematical treatment is simple, as shown by a numerical investigation for multiple stages. The transient backflow cell model (BCM), based on the first principle of mass balances, was developed and implemented in MATLAB®. Meanwhile, the transient cell model (CM) was implemented as a reference to simulate the absorptive reactor when assuming no backflow and, therefore, no absorption effect. The sensitivity on the number of cells, as one parameter used in the model, is also considered.

## 2. EXPERIMENTAL SECTION

The synthesis of Cs-Ru/MgO followed the incipient wetness impregnation procedure. The effect of the absorbent on ammonia production was tested in a fixed-bed continuous-flow reactor using Cs-Ru/MgO catalyst. A glass reactor tube (cross-sectional area  $5.024 \times 10^{-5} \text{ m}^2$ ) was packed with 0.1 g Cs-Ru/MgO catalyst and 1 g absorbent ( $\text{MgCl}_2$ ) with quartz wool in between. A 3:1

mixture of H<sub>2</sub> and N<sub>2</sub> was passed through the reactor at a pressure of 10 bar and a weight hourly space velocity of 36000 mL g<sup>-1</sup> h<sup>-1</sup>. The reactor was equipped with three independent furnaces that allowed different sections to be heated to different temperatures. Each furnace was controlled by an independent thermocouple. The middle thermocouple was directly inserted into the catalyst bed and was set to synthesis temperature. During the reaction step, the catalyst was at 400 °C and the absorbent at 150 °C. Finally, any ammonia absorbed on the absorbent was desorbed by heating the absorbent to 400°C under flowing nitrogen at ambient pressure. The reaction rate of the catalyst was 7884 μmolg<sup>-1</sup>h<sup>-1</sup> with the absorbent loaded behind the catalyst and 5082 μmolg<sup>-1</sup>h<sup>-1</sup> without the absorbent.

### 3. MATHEMATICAL MODEL

The backflow process leads to two possible mechanisms that affect the single-pass conversion of the reaction: increased residence time and ammonia removal due to the absorption effect. In this section, we identify and compare the extent of each mechanism via two models: the transient BCM and the transient CM for the absorptive reactor (catalyst bed adjacent to an absorbent bed) and the conventional reactor (catalyst bed only). Figure 1 represents a schematic of the transient BCM in an absorptive ammonia synthesis reactor.

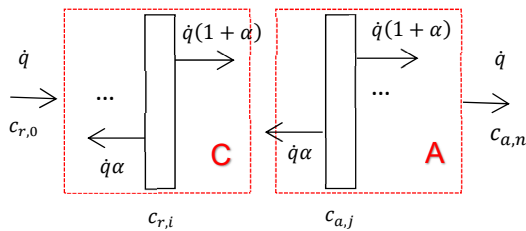


Figure 1: Schematic Diagram of the BCM in the Ammonia Synthesis Absorptive Reactor ('C' refers to the catalyst bed while 'A' refers to the absorbent bed).

The BCM hypothesises a stable unchanging backflow, expressed by  $\dot{q}\alpha$  between each cell, to characterise the backflow mechanism in the reactor. There are  $m$  cells in the catalyst bed and  $n$  cells in the absorbent bed. In this case, the backflow ratio  $\alpha$  represents the backflow amount. An  $\alpha$  of zero is the plug flow limit; while when  $\alpha$  equals infinity, the equations model approaches the stirred tank limit. The assumptions that govern the development of transient BCM are as follows:

- Due to symmetry in the radial direction of the packed bed, the governing equations are independent of this direction. Thus, a one-dimensional model is considered.
- Constant actual (interstitial) velocity is assumed.
- Isothermal condition is assumed in the catalyst bed.
- Each cell is assumed to be perfectly mixed.

- The volumetric flow of gas is assumed to be constant throughout the axial position of the bed.
- Instantaneous equilibrium of the solute in the bulk fluid with the absorbate.
- Catalyst/absorbent packing has uniform voidage and particle size.

According to these assumptions, the differential equations of the transient BCM were developed along the axial direction based on the principle of mass balance. The mathematical relationship for these cells can be written as follows. For the first cell of the catalyst bed:

$$V_{r,f} \cdot \frac{dc_{r,1}}{dt} = \dot{q}c_{r,0} + \dot{q}\alpha c_{r,2} - \dot{q}(1+\alpha)c_{r,1} + V_{r,c} \cdot R_r \quad (1)$$

Similarly, the modelling equations for the intermediate cells of the catalyst bed ( $i = 2, 3, \dots, m-1$ ) can also be written.

$$V_{r,f} \cdot \frac{dc_{r,i}}{dt} = \dot{q}(1+\alpha)c_{r,i-1} + \dot{q}\alpha c_{r,i+1} - \dot{q}(1+2\alpha)c_{r,i} + V_{r,c} \cdot R_r \quad (2)$$

For the last cell of the catalyst bed:

$$V_{r,f} \cdot \frac{dc_{r,m}}{dt} = \dot{q}(1+\alpha)c_{r,m-1} + \dot{q}\alpha c_{r,m+1} - \dot{q}(1+2\alpha)c_{r,m} + V_{r,c} \cdot R_r \quad (3)$$

For the first cell of the absorbent bed:

$$V_{a,f} \cdot \frac{dc_{a,1}}{dt} = \dot{q}(1+\alpha)c_{a,0} + \dot{q}\alpha c_{a,2} - \dot{q}(1+2\alpha)c_{a,1} + V_{a,c} \cdot R_a \quad (4)$$

For the intermediate cells of the absorbent bed,  $j = 2, 3, \dots, n-1$ :

$$V_{a,f} \cdot \frac{dc_{a,j}}{dt} = \dot{q}(1+\alpha)c_{a,j-1} + \dot{q}\alpha c_{a,j+1} - \dot{q}(1+2\alpha)c_{a,j} + V_{a,c} \cdot R_a \quad (5)$$

For the last cell of the absorbent bed:

$$V_{a,f} \cdot \frac{dc_{a,n}}{dt} = \dot{q}(1+\alpha)c_{a,n-1} - \dot{q}(1+\alpha)c_{a,n} + V_{a,c} \cdot R_a \quad (6)$$

where  $\dot{q}$  is the inlet gas flowrate (in m<sup>3</sup>/s) and  $\dot{q}\alpha$  is the backflow. The forward flow increases to  $\dot{q}(1+\alpha)$ . The mass balances for hydrogen, nitrogen, and ammonia are respectively calculated by Equations (1) through (6), with different reaction rates and absorption rates.



Dyson and Simon modified the Temkin expression to calculate the intrinsic rate of reaction in  $\text{kmol/m}^3$  (Dyson & Simon 1968), given in Equation (7):

$$R_{r,NH_3} = 2k_2 \left[ K_a^2 a_{N_2} \left( \frac{a_{H_2}}{a_{NH_3}} \right)^\beta - \left( \frac{a_{NH_3}}{a_{H_2}} \right)^{1-\beta} \right] \quad (7)$$

The parameter  $\beta$  is a constant, taking values between 0.5 and 0.75, and the expression of the activity of the components,  $a_k$ , is shown in Equation (8):

$$a_k = \frac{f_k}{P^0} = \frac{\phi_k y_k P}{P^0} \quad (8)$$

$P^0$  is the reference pressure and is assumed to be atmospheric (1 bar). Gillespie and Beattie calculated the equilibrium constant,  $K_a$ , and proposed the correlation in Equation (9) (Gillespie and Beattie 1930):

$$\log_{10} K_a = -2.691122 \log_{10}(T) - 5.519265 \times 10^{-5} T + 1.848863 \times 10^{-7} T^2 + \frac{2001.6}{T} + 2.689 \quad (9)$$

In turn,  $k_2$  is estimated by the Arrhenius equation (Gillespie & Beattie 1930):

$$k_2 = 8.849 \times 10^{14} e^{\left( \frac{-40765}{1.9877 T} \right)} \quad (10)$$

The absorbents only absorb ammonia. Therefore, the absorption rate for hydrogen and nitrogen is zero.

In this mathematical model, the absorbent bed is first assumed to be in the non-saturated condition. Then, the saturation condition of the absorbent is additionally considered to determine the influence of the absorbent amount. While a detailed mechanism for absorption is not known, kinetic experiments conducted previously (Smith et al. 2018) allow the fitting of an analytical equation:

$$r_{abs} = k_{a1} e^{k_{a2}(P_{NH_3} - P_{NH_3}^+)} \text{ when } (P_{NH_3} - P_{NH_3}^+) < 1 \text{ bar} \quad (11)$$

$$r_{abs} = k_{a3} (P_{NH_3} - P_{NH_3}^+)^2 \text{ when } (P_{NH_3} - P_{NH_3}^+) > 1 \text{ bar}$$

where  $r_{abs}$  is the rate in micromoles per second per gram of absorbent,  $k_{a1-a3}$  are experimentally determined constants (Smith et al. 2019),  $P_{NH_3}$  is the pressure of ammonia in kPa, and  $P_{NH_3}^+$  is the equilibrium pressure of ammonia at a given temperature in kPa, as illustrated in (Smith et al. 2019). Based on these analytical equations, the temperature profile of the absorbent bed could be

ignored as the parameters and equations were assumed to be independent of temperature-irrelevant.

The governing equation for the conventional reactor in BCM is Equation (12). To some extent, the existing of backflow can be assumed as that the flow goes through the catalyst bed with longer residence time. Therefore, the cell number of the catalyst bed is  $m'$ , which is higher than  $m$ .

$$V_{r,f} \cdot \frac{dc_{r,i'}}{dt} = \dot{q}c_{r,i'-1} - \dot{q}c_{r,i'} + V_{r,c} \cdot R_r \quad (12)$$

In the CM of the absorptive and the conventional reactors, the reactants flow in one direction. The modelling equations for the cells of the catalyst bed were consistent with Equations (1) to (6) but ignored the backflow.

#### 4. SIMULATION AND DISCUSSION

The governing equations of the transient BCM and transient CM are ordinary differential equations which were solved using the ode15s solver from MATLAB®. The numerical results and discussion are presented in this section. Nitrogen is taken as the reference to calculate the reaction conversion.

##### 4.1 The Conventional Reactor

Figure 2 depicts the concentration of these three components in the catalyst bed at different times (arrow points in the direction of an increasing number of cells). The steady-state concentration of hydrogen and nitrogen for the downstream position is lower than upstream positions as the reaction progresses.

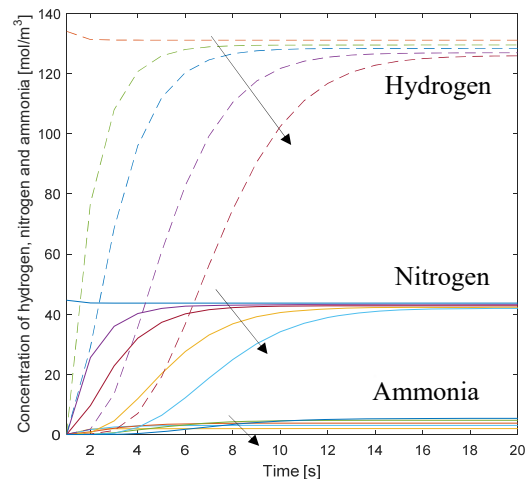


Figure 2: Concentration of hydrogen, nitrogen, and ammonia in the conventional reactor.

As illustrated in Figure 3, the concentration of ammonia increases from the first cell to the end as a result of the superposition of the produced ammonia (arrow points in the direction of an increasing number of cells). The

conversion in the catalyst bed is 6.25% according to the simulation results, which depend on the parameters used in the model, such as the length of the catalyst bed, temperature, and pressure.

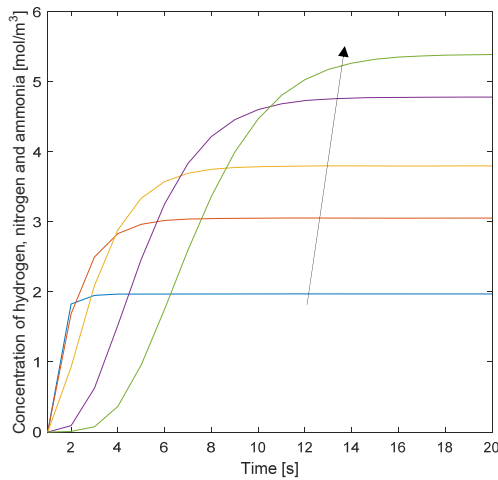


Figure 3: Concentration of ammonia in the conventional reactor.

Therefore, the reaction equilibrium has not been reached, as the thermodynamic equilibrium conversion of ammonia synthesis from  $H_2: N_2 = 3:1$  at  $T = 400\text{ °C}$  and  $P = 10\text{ bar}$  is 7.45% (Ogura et al. 2018). Then, the effect of the catalyst bed length on the reaction conversion was considered here. As demonstrated in Figure 4, the conversion of the ammonia synthesis rises sharply at first with the increasing length of the catalyst bed, then reaches a stable value: the equilibrium point of this reversible reaction. This balance cannot be broken unless the reaction conditions change. In Figure 4, the ammonia synthesis reaches the equilibrium when the catalyst bed is longer than 0.04 m. The synthesis conversion is independent of the residence time if the equilibrium has been reached.

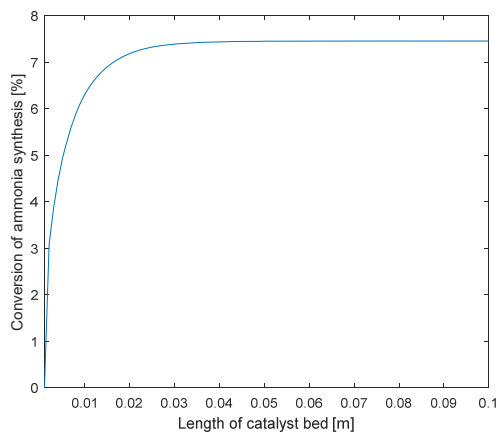


Figure 4: Variation of the conversion of ammonia synthesis with the length of the catalyst bed.

Figure 5 presents the conversion of ammonia synthesis under various backflow extents (0% to 100%) when the

length of the catalyst bed is 0.01 m. An increase in ammonia production conversion from 6.25% to 7.25% can be observed. In the transient BCM of the conventional reactor, the extent of backflow was assumed to be related to the number of cells in the catalyst bed, which replace the increase in residence time. When the extent of backflow is 100% the catalyst bed is assumed to be extended by about 10 cells, which means the flow has a longer residence time for reaction.

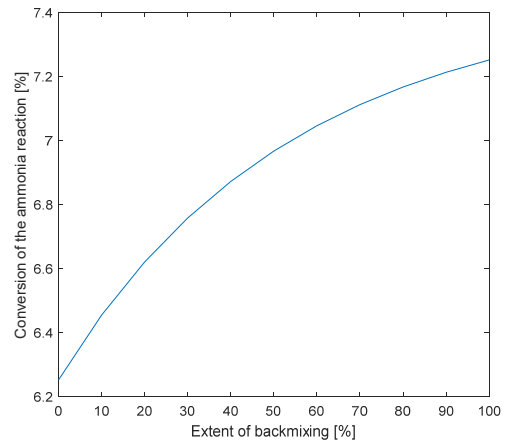


Figure 5: Effect of backflow on the ammonia synthesis conversion.

The number of cells is an important parameter in the BCM and CM. Therefore, it is necessary to analyse the sensitivity of the number of cells at a fixed-length catalyst bed, which is displayed in Figure 6 (arrow points in the direction of an increasing bed length). The cell number sensitivity study is a significant aspect to avoid inaccurate results due to an ill-conditioned cell number. In Figure 6, the cell number should be larger than 8 when the length of the catalyst bed is longer than the equilibrium length, and then the results stop changing with cell number. When the length of the catalyst bed is shorter than 0.04 m, increasing the number of cells will slightly increase the reaction conversion. Therefore, in this case, a cell number of 12 is taken to minimise the deviation of cell number dependence.

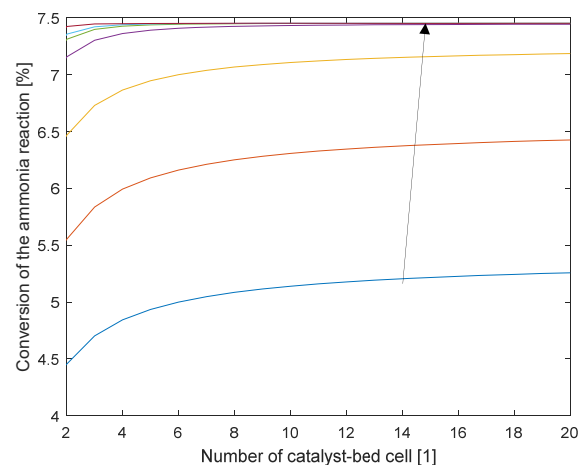


Figure 6: Sensitivity on the number of cells in the catalyst bed.

#### 4.2 The Absorptive Reactor

Based on the results in Section 4.1 for the simulation of the absorptive reactor, the length of the catalyst bed was fixed to 0.1 m to ensure that the reaction equilibrium was reached and the cell number was set to be 8 to decrease computation time. The single-pass conversion of the absorptive reactor is 7.45%, calculated by the transient CM, which is the same as the conversion of the conventional reactor (0.1 m length). Therefore, the absorption does not affect the ammonia synthesis conversion when using the transient CM, which is not consistent with the experimental results.

Based on the experimental results, an absorption bed following the catalyst bed has an impact on the single-pass reaction conversion. Therefore, the transient CM does not apply to the absorptive reactor and the simulation can be switched to the transient BCM with a fixed 50% backflow ratio. The solid lines in Figures 7 and 8 display the concentration profiles of nitrogen and ammonia. Unlike the transient CM results (dashed lines), the ammonia concentration decreases at the end of the catalyst bed, when the flow has yet to reach to absorbent bed. The nitrogen concentration has a second drop at the junction of the catalyst bed and absorbent bed, shown in the red square of Figure 7.

The comparison of the transient BCM and CM simulations is illustrated in Figures 7 to 8 for a 50% backflow ratio. The arrows point in the direction of increasing time. As illustrated in Figures 7 and 8, the dashed curves and solid curves are noticeably different, which means that simulated results were influenced greatly by the backflow setting. Under the same cell number and same time conditions, a lower nitrogen concentration with higher ammonia concentration is revealed for the absorptive reactor with the BCM simulation in comparison with that in the CM simulation. This implies that the single-pass synthesis conversion in the BCM simulation is higher than the one predicted by the CM simulation. Undoubtedly, the transient BCM matched experimental results better than transient CM.

Based on the BCM simulation, the absorption capacity of the absorptive reactor, which promotes the reaction conversion, can be further analysed. In this BCM simulation, the equilibrium of the reaction in the catalyst bed has been reached as its length is set at 0.1 m with 8 cells. Therefore, the single-pass conversion of ammonia synthesis does not change when there is the only backflow without absorption or increased residence time, based on the results of Section 4.1. Therefore, the reason for improved conversion is due to the absorption capacity of the absorptive reactor, which speeds up the reaction rate and drives the reaction equilibrium as the partial pressure of ammonia decreases. Then, the backflow ratio,

length and cell number of absorbent bed, as well as the saturation degree of absorbent bed, will be discussed.

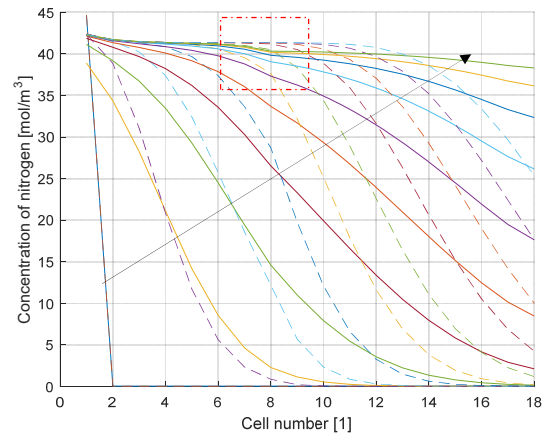


Figure 7: Concentration of nitrogen with respect to the position in the absorptive reactor.

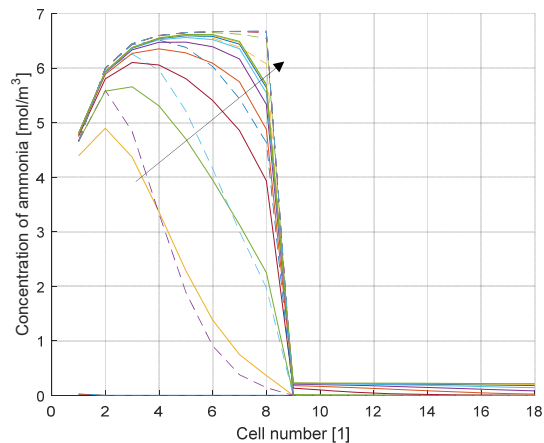


Figure 8: Concentration of ammonia with respect to the position in the absorptive reactor.

Figure 9 reveals a relationship between the synthesis conversion of the absorptive reactor and the backflow ratio. The reaction conversion has a linear growth as the backflow ratio increases when the absorbent bed has not been saturated. The reaction conversion increases from 7.45% (the equilibrium conversion) to 11.2% in Figure 9. Therefore, the absorption in the absorptive reactor can drive the synthesis equilibrium in the right direction when the thermodynamic equilibrium of the reaction has already been reached.

Then, the length and cell number of the absorbent bed were considered for their impact on the conversion improvement. Based on the simulation results, it is illustrated that increasing infinitely the length or the cell number of the absorbent bed does not affect the synthesis conversion when the absorbent bed was assumed to be unsaturated.

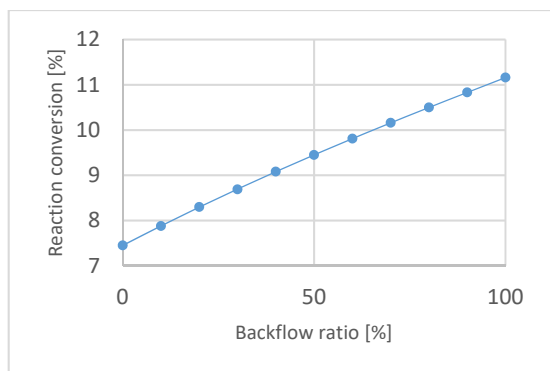


Figure 9: Relationship between reaction conversion and backflow ratio in the absorptive reactor.

However, in reality, the absorbent bed is gradually saturated as the synthesis take place, and the effect of absorption on the synthesis conversion begins to wane as the absorbent pellets stop absorbing ammonia. For example, Figure 10 displays a relationship between the absorptive reactor conversion and the saturated cells in the absorbent bed. In this case, the backflow ratio is set at 50% with a fixed-absorbent-bed length of 0.1 m. A saturated cell number equal to six means that the first six cells of the absorbent bed have been saturated and the absorption occurs at the third cell. Figure 10 reveals that the saturation state of only the first six cells in the absorbent bed can affect the synthesis conversion. Once the first six cells are saturated, the conversion no longer increases, and there is no difference between an absorptive reactor and a conventional reactor.

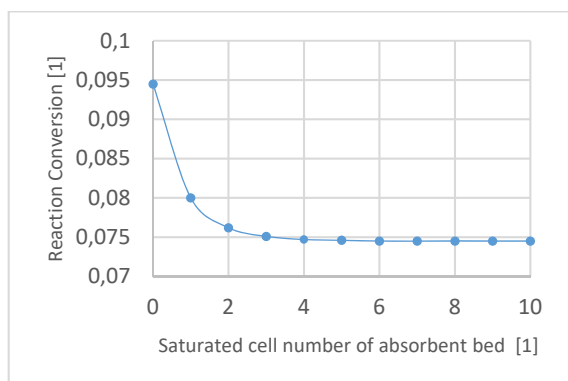


Figure 10: Reaction conversion of ammonia synthesis with a gradually saturated absorbent bed.

When the ammonia synthesis in the catalyst bed has not reached the thermodynamic equilibrium, the results indicate that, under the kinetically controlled regime, both the longer residence time and the absorption of ammonia can yield higher ammonia conversion to approach the thermodynamic value as shown in our model. If the packing after the catalyst bed contains solid pellets but no absorbent, the reaction conversion can be improved only by the residence time. However, once the equilibrium is reached, the increase of the reaction conversion is due to the absorption. The backflow ratio

can be regarded as the scale of the fluid disorder in the absorptive reactor, which will directly relate the impact extent of absorption on synthesis. If the backflow ratio is zero, the decrease in ammonia concentration or partial pressure due to absorption will not affect the synthesis reaction. Furthermore, when the backflow ratio is one hundred, the reactants with the reduced partial pressure of ammonia (ammonia was absorbed) will flow to upstream cells, speeding up the reaction and driving the equilibrium.

It can be summarized that the reaction conversion is entirely independent of the cell number of the absorbent bed. Once the absorbent bed length is long enough, e.g. at least 0.1 m, the reaction conversion will not increase even if the absorbent bed is longer. Moreover, if the first several cells (about 0.06 m) of the absorbent bed are saturated, the absorption will not affect the synthesis conversion any more.

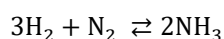
The reaction conversions for experimental results were 1.09% and 0.71%. These were much lower than the simulation conversion of a stoichiometric feed, which was due to the activity of the catalyst being calculated based on an iron catalyst (Dyson and Simon 1968) in the simulation, not on the Ru catalyst that was used in the experiment as the lack of rate expression for Ru catalyst. Furthermore, the data at mild conditions were not a good comparison for the reaction rates under ideal simulation conditions. However, as the backflow effects on the single-pass conversion in the simulation have the same increasing trend as in reality, this model of the ammonia absorptive reactor can, to some extent, explain the experimental results.

There are potential limitations to this simulation. In reality, both ammonia synthesis and absorption are highly exothermic. However, in this simulation, we assumed that the temperature of the catalyst bed was constant at 400 °C, while it declined linearly at the beginning of the absorbent bed from 400 °C to 150 °C. Also, the ammonia synthesis reaction is not equimolar (with 4 mol of reactants being converted to 2 mol of products), and absorption will only remove ammonia from the gas phase. In this simulation, the single-pass conversion of the reaction is quite low at about 5 – 7 %. Therefore, the volume change during the reaction and absorption processes was low enough to be ignored in this paper. Nevertheless, these are future simulation we would like to perform.

## 5. CONCLUSION

In this paper, a transient BCM was performed to simulate an absorptive reactor for ammonia synthesis. Meanwhile, the transient CM was taken as a reference to identify what role the absorbent plays in the enhancement of synthesis conversion. Modelling results indicate that the absorbent packed adjacently after the catalyst significantly affects the single-pass synthesis conversion through increased

residence time and faster reaction rate, driving equilibrium to the right. The backflow increases the residence time, while the faster reaction rate and equilibrium-shift are due to the reduced ammonia partial pressure (absorbed in the absorbent). The bench-scale absorptive ammonia synthesis reactor was operated to obtain experimental data to underline this point. Furthermore, the absorption has more impact on the conversion improvement than the increased residence time, as the proportional increase in Figure 9 is more significant than that in Figure 5. These two models could demonstrate that backflow in the absorptive reactor promotes the ammonia reaction conversion via two mechanisms: longer residence time for reaction and a faster reaction rate, driving the equilibrium of the reaction to the right (reversible reaction shown as below).



## REFERENCES

- Appl, M., 1999. *Ammonia: principles and industrial practice*, Weinheim: Wiley-VCH.
- Bruce, A. and Faunce, T., 2015. Sustainable fuel, food, fertilizer and ecosystems through a global artificial photosynthetic system: Overcoming anticompetitive barriers. *Interface Focus*, 5(3), pp.1–9.
- Dyson, D.C. and Simon, J.M., 1968. A kinetic expression with diffusion correction for ammonia synthesis on industrial catalyst. *Industrial & Engineering Chemistry Fundamentals*, 7(4), pp.605–610.
- Gillespie, L.J. and Beattie, J.A., 1930. The Thermodynamic Treatment of Chemical Equilibria in Systems Composed of Real Gases. I. An Approximate Equation for the Mass Action Function Applied to the Existing Data on the Haber Equilibrium. *Physical Review*, 36(4), pp.743–753.
- Huberty, M.S. et al., 2012. Ammonia absorption at haber process conditions. *AIChE Journal*, 58(11), pp.3526–3532.
- McSwain, C. V and Durbin, L.D., 1966. The Backflow-Cell Model for Continuous Two-Phase Nonlinear Mass-Transfer Operations Including Nonlinear Axial Holdup and Mixing Effects. *Separation Science*, 1(6), pp.677–700.
- Ogura, Y. et al., 2018. Efficient ammonia synthesis over a Ru/La<sub>0.5</sub>Ce<sub>0.5</sub>O<sub>1.75</sub> catalyst pre-reduced at high temperature. *Chem. Sci.*, 9(8), pp.2230–2237.
- Sinkule, J., Hlaváček, V. and Votruba, J., 1976. Modeling of chemical reactors-XXXI. The one-phase backflow cell model used for simulation of tubular adiabatic reactors. *Chemical Engineering Science*, 31(1), pp.31–36.
- Smith, C. et al., 2018. Rates of Ammonia Absorption and Release in Calcium Chloride. *ACS Sustainable Chemistry & Engineering*, 6(9), pp.11827–11835.
- Smith, C., McCormick, A. V and Cussler, E.L., 2019. Optimizing the Conditions for Ammonia Production Using Absorption. *ACS Sustainable Chemistry and Engineering*, 7(4), pp.4019–4029.
- Wagner, K. et al., 2017. Column absorption for reproducible cyclic separation in small scale ammonia synthesis. *AIChE Journal*, 63(7), pp.3058–3068.
- Zhu, H. et al., 2009. Large-Scale Synthesis of MgCl<sub>2</sub> · 6NH<sub>3</sub> as an Ammonia Storage Material. , pp.5317–5320.

## AUTHOR BIOGRAPHIES

**WENCHAN QI** was born in Shandong, China and went to the University of Birmingham, where she studied chemical engineering and obtained her MEng degree in 2016. She then moved to the University of Oxford for DPhil degree study, and she is now working on a research of the optimisation of the Haber-Bosch reactor. Her e-mail address is wenchan.qi@eng.ox.ac.uk.

**RENE BANARES-ALCANTARA** has an MEng from UNAM (Mexico), and an MSc and PhD from Carnegie Mellon University (USA), all of these degrees in Chemical Engineering. He has worked in the Department of Engineering Science at Oxford since 2003 and is a Fellow of New College. His research interests are in the area of Process Systems Engineering, mainly process design, synthesis and simulation. Since 2014 he has been involved in projects related to long-term (chemical) storage of renewable energy and the production of ‘green’ ammonia.

# Improved Decentral Task Allocation for Autonomous Guided Vehicle Systems based on Karis Pro

Maximilian Selmair  
BMW Group  
80788 Munich, Germany  
maximilian.selmair@bmw.de

Klaus-Jürgen Meier  
University of Applied Sciences Munich  
80335 Munich, Germany  
klaus-juergen.meier@hm.edu

**Keywords**—Task Allocation; AGV; Karis Pro

**Abstract**—In this paper, we extended an existing decentralised method for allocating tasks to AGVs, by additionally considering vehicles which already are assigned to a task. This was achieved by also considering the opportunity costs arising from a vehicle passing a current task to another vehicle and subsequently accepting a new task. This loosened restriction is enabling the vehicle fleet for a higher flexibility, which can be used for improving the efficiency of the overall system. By means of simulation, our findings confirm the notion that our extended method – namely Karis Pro+ (KP+) – leads to lower traffic density and higher flexibility, both of which are important KPIs for large-scale transport vehicle systems.

## LIST OF ABBREVIATIONS

**AGV** Automated Guided Vehicle  
**BMBF** Bundesministerium für Bildung und Forschung  
(English: German Federal Ministry of Education and Research)  
**FCFS** first-come-first-serve  
**KIT** Karlsruhe Institute of Technology  
**KP** Karis Pro  
**KP+** Karis Pro+  
**KPI** Key Performance Indicator  
**MRPD** Multi-robot Task Allocation for Pickup and Delivery  
**MRS** Multi-robot System  
**MRTA** Multi-robot Task Allocation  
**TeSSI** Temporal Sequential Single-item Auction  
 $V_I$  idle vehicles  
 $V_P$  vehicles in pickup  
 $V_D$  vehicles in delivery  
 $R_A$  assigned transportation requests  
 $R_U$  unassigned transportation requests

## I. INTRODUCTION

Greater efficiency, higher productivity and more flexibility are some of the key objectives that govern today's production operations (Dorri et al. 2018). Consequently, these objectives are the focus of strategic efforts regardless of company's industrial branch. Production and logistics have been identified as the two areas in manufacturing companies that are especially subject to meet these demands, which entails the necessity for continuous innovation and improvements (Dorri et al.

2018). The main driving forces behind these objectives and the associated requirements are the competitive pressure of the market as well as customer demands for greater customisation and individualisation (Leupold et al. 2018; Jazdi 2014). In order to meet these requirements and to maintain a competitive edge in its industry, new technologies for handling material are in demand. Conventional material flow systems, seemingly rigid in terms of layout and throughput, are therefore continuously becoming less suitable to meet industry standards. The constantly rising demand for flexibility drives the development of autonomous floor-bound vehicle systems. One of these vehicle systems is the result of the research project Karis Pro, introduced by Colling et al. (2016) as a "modular, decentralised controlled automated guided vehicle system". This task allocation system harnesses both, a decentralised job creation system and a decentralised job allocation system. Due to its flexible application, the developers state this system can be employed with ease by the customer and without the need for infrastructural adjustments. The decentralised control and the autonomous nature of each vehicle forms a reliable system that caters to the requirements of modern industrial processes (Colling et al. 2016).

Karis Pro (KP) has been utilised in two industrial settings under real-life conditions. An industrial pilot study, set in the production lines, was successfully performed in corporation with Bosch and Audi (see Figure 1). The use cases of autonomous transport systems in this environment are driven by the principles of lean production, in order to reduce the produced and transported lot sizes as well as to increase the frequency of arrivals and departures without having to increase the number of required personnel (Trenkle 2016).

Particularly large-scale vehicle systems depend on highly efficient organisation to avoid traffic jams and delays (Versteegt and Verbraeck 2002). A means of achieving this may be to allocate scarce resources, such as the number of available vehicles, as efficiently as possible. Our developed method aims to improve the established method of Colling et al. (2016) to enhance the performance and efficiency significantly.

This paper is divided into the following sections. Section II provides a review of the relevant literature, followed by an introduction of the extended version of Karis Pro, namely Karis Pro+ and an outline of the associated development ratio-



Fig. 1. Karis Pro AGV in an Industrial Use Case; Illustration from [amplify.pepperl-fuchs.com/de/inhalte/121/wo-die-dinge-ins-schwaermen-geraten](http://amplify.pepperl-fuchs.com/de/inhalte/121/wo-die-dinge-ins-schwaermen-geraten)

nale. A description of a subsequent simulation study to validate the improvements in Section IV precedes the conclusion in Section V.

## II. RELATED LITERATURE

In academia, the issue of task allocation to agents is known as the *Multi-robot Task Allocation* (MRTA) problem or, more specific to the use-case of this study, the *Multi-robot Task Allocation for Pickup and Delivery* (MRPD) problem. The MRPD defines an extension of the widely studied MRTA problem which assumes each task as a single location to visit (Heap and Pagnucco 2013). By definition, a *Multi-robot System* (MRS) describes a group of agents that are assigned to perform a collective behaviour (Khamis, Hussein, et al. 2015). Due to its high industrial academic relevance, one of the major topics of research in the area is the multi-robot task allocation problem. That is the problem of dividing tasks across robots so that some objective of interest is optimised (Nunes, Manner, et al. 2017). Referring to Gerkey and Mataric (2004), this particular MRPD problem can be described as ST-SR-IA (single-task robots, single-robot tasks, instantaneous assignment).

Contributions around MRTA are utilised in many different industries, such as for fleets of cooperated unmanned aerial vehicles (Moon et al. 2015; Bellingham et al. 2003; Innocenti et al. 2010), maritime industry (Bădică et al. 2018), manufacturing automation (McIntire et al. 2016), robots for refinery inspections (Liu and Kroll 2012) and areas where it is either too expensive or dangerous for humans to work (Nagarajan and Thondiyath 2013; Khamis and ElGindy 2012). In order to find a suitable solution for the task allocation problem, the literature proposes several different concepts. The review of the literature indicates that the most used and state-of-the-art solution for a large-scale decentralised system in order to achieve satisfactory performance is the agent-based auction mechanism. Auctions, in any form, have been used in societies throughout history to allocate scarce resources among interested parties (Gerkey and Mataric 2002). Consequently, auctions have received a great deal of attention from economists, and a rich body of theory exists regarding their properties, especially their ability to deal with uncertainty (Akerlof 1970; Holt 1979).

For common industrial applications, agents such as robots

or vehicles, bids on tasks by calculating the amount of effort needed to perform them. In most cases, the effort depends on the distance between the agent location and the task location plus any additional cost for doing the task itself, such as resources consumed (e. g. time spent) in doing the task (Nunes and Gini 2015). The agents act as bidders and stations as sellers. Thereupon, the agents submit their bids and, depending on this, receive an acceptance or refusal (Schwarz et al. 2013).

The smart coordination of groups of robots is an important topic in this context, as significant amount of resources can be saved when robots work together. While this does not necessarily mean that the robots conduct their tasks together in teams, but they do adjust their actions to each other (Gerkey and Mataric 2002).

One should also pay attention to spatial and temporal constraints, as many tasks have to be executed in a specified time window in industrial practice, as well. For this reason, the *Temporal Sequential Single-item Auction* (TeSSI) algorithm was developed (Nunes and Gini 2015; Nunes, Manner, et al. 2017). In addition, Bădică et al. (2018) proposed an “intelligent” freight broker agent which uses a mathematical optimisation service to compute the optimal schedule of vehicles fulfilling customer requirements.

Using heuristic-based task allocation algorithms is another strategy to solve such a task allocation problem (Jennings et al. 2001; Windelinckx and Strens 2004; Nagarajan and Thondiyath 2013). Such static approaches usually require that the announced tasks and the available vehicles are known prior to the actual calculation and scheduling phase. As this is not the case for a decentralised control setting, these approaches are of no interest for this research (Schwarz 2014).

Summarised, agent-based methods like negotiations (e. g. auctions) are an appropriate and established method for proactive allocation of tasks to vehicles (Lagoudakis et al. 2004; Tovey et al. 2005). One showcase project for a decentral task allocation will be presented in detail in the next section. Furthermore, Karis Pro will be enabled for the idea of different traveller assignment strategies, presented by Hyland and Mahmassani (2018).

## III. KARIS PRO AND THE TRAVELLER ASSIGNMENT STRATEGIES

Karis Pro is a research project of the Karlsruhe Institute of Technology (KIT), which was carried out between 2013 and 2016 with the support of the German Federal Ministry of Education and Research (BMBF).

As opposed to the centralised methods, the process of allocating tasks to vehicles within the Karis Pro system takes place without the involvement of a central control unit. Instead, the vehicles carry out auctions in order to determine which vehicle should handle which task (Trenkle 2016). After a task is generated, it is automatically entered into a list that ranks all tasks according to a specific priority. The tasks with the highest priority can be found at the top of the list. As Colling et al. (2016) has not specified which priority was used in their approach, different criteria for the priority rule could be selected depending on the underlying specifications. Possible examples of these priorities are:

- Date of task generation, i. e. according to the first-come, first-served rule, the oldest tasks have the highest priority
- Deadline when a task has to be fulfilled, i. e. tasks closer to the deadline have a higher priority
- Distance between origin (source) and destination (sink), i. e. shorter distances have a higher priority or vice versa
- Pre-determined rules, i. e. certain source-sink relations are more critical and have thus a higher priority

The tasks are assigned to vehicles just before their execution, which means that each vehicle receives only one task at a time. The task allocation process begins when one of two events takes place. Either a vehicle has just finished its task and there are tasks left on the list, or a vehicle is idle and a new task appears on the list. In both instances the vehicle chooses the task with the highest priority and begins the auction process. During the first step of the auction, the vehicle that started the auction calculates its own bid. Subsequently, it contacts all other vehicles and asks them to submit their bids. These vehicles can be categorised as follows:

- 1) *Idle*: Without a task
- 2) *Pickup*: Vehicle is currently driving to a source to pick up a task. This can also be described as an "empty drive" as it is not adding value
- 3) *Delivery*, without reservation: Vehicle is currently transporting a task to its sink and is not reserved for a subsequent task.
- 4) *Delivery*, with reservation: Vehicle is currently transporting a task to its sink and is reserved for a subsequent task.

Under the constraints of *Karis Pro*, vehicles from Category 1 (*Idle*) and Category 3 (*Delivery*, without reservation) will respond with a bid to the auctioning vehicle. The bid itself is given in number of seconds and is either appraised or calculated by adding the following values:

- 1) The expected remaining time required to fulfil the current task (only for vehicles of Category 3)
- 2) The expected time to reach the source of the new task (for vehicles of Categories 1 and 3)

Therefore, the bid is the amount of time it would take each vehicle to reach the source of the new task. After having received all bids, the auctioning vehicle determines the most efficient vehicle, which can lead to either one of two outcomes. The vehicle itself is the most suitable for the task (i. e. the lowest bid) or another vehicle is more favourable. In the first case, the auctioning vehicle immediately begins to execute the task and informs the other vehicles that it was awarded with the task. In the second case, the auctioning vehicle informs the winner that it has won the auction. Subsequently, it chooses the next task on the list and commences a new auction. The auctioning vehicle continues to repeat the auctioning process until it either wins an auction, there are no more tasks left on the list or no other vehicles remain to participate in auctions.

The upcoming presented contribution deals with assignment strategies, which are a substantial part of task allocation methods (Nunes, Manner, et al. 2017). Hyland and Mahmassani (2018) have compared different assignment strategies for a share-use autonomous vehicle mobility service. In this setting, the vehicles are controlled by a central unit. This information

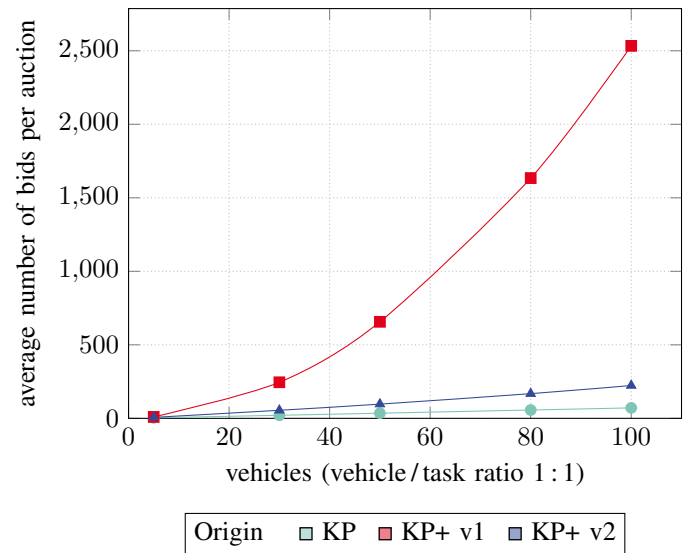


Fig. 2. Average Number of Bids per Auction (5,000 Samples)

hub provides a direct origin-to-destination service to travellers who request a ride, or, if one were to apply this to the environment of our research, for a source-to-sink service of a load carrier.

Table I summarises the different strategies compared by Hyland and Mahmassani (2018). The first two assignment strategies are basic first-come-first-serve (FCFS) strategies with different objectives, to either minimise waiting time or travelling distance. Here, unassigned travellers ( $R_U$ ) are assigned to idle vehicles ( $V_I$ ). The other four optimisation strategies consider different combinations of idle vehicles, vehicles in pickup ( $V_P$ ) and vehicles in delivery ( $V_D$ ) as well as the variables unassigned or assigned travellers ( $R_A$ ). In order to generate the most efficient outcome in light of the given objectives, a mathematical solver is used to identify the most suitable allocation. The optimisation-based strategies, especially those that involve the reassignment of already assigned travellers ( $R_A$ ), significantly outperform both FCFS assignment strategies.

More specifically, the results of their simulation study indicate that Strategy 6 unambiguously outperforms all other strategies in terms of empty fleet miles in all of the examined scenarios. In terms of traveller waiting times, Strategy 6 is deemed to be more efficient when the fleet size is small relative to the demand rate. For large fleet sizes like 200+ vehicles, Strategy 3 and Strategy 4 outperform Strategy 5 and Strategy 6 in terms of traveller waiting times. It is proposed that Strategy 5 in Table I corresponds to the current *Karis Pro* method.

#### IV. METHODOLOGY

The basic methodology to achieve this research's objective, is to compare the existing method of Colling et al. (2016) with our extended version, namely *Karis Pro+*. This comparison is accomplished by means of a simulation study in which the agent-based modelling paradigm is used.

The software *AnyLogic* is chosen for this purpose, as it supports the paradigms of agent-based modelling and thus



TABLE I. OVERVIEW OF TRAVELLER ASSIGNMENT STRATEGIES (HYLAND AND MAHMASSANI 2018)

	Strategy	Travelers ( $R'$ )	Vehicles ( $V'$ )	Sequential/Simultaneous	Traveler Reassignment?	En-Route Drop-off
1	First-Come-Frist-Serve	$R_U$	$V_I$	Sequential	No	No
2	First-Come-First-Serve	$R_U$	$V_I$	Sequential	No	No
3	Optimization	$R_U$	$V_I$	Simultaneous	No	No
4	Optimization	$R_U \cup R_A$	$V_I \cup V_P$	Simultaneous	Yes	No
5	Optimization	$R_U$	$V_I \cup V_D$	Simultaneous	No	Yes
6	Optimization	$R_U \cup R_A$	$V_I \cup V_P \cup V_D$	Simultaneous	Yes	Yes

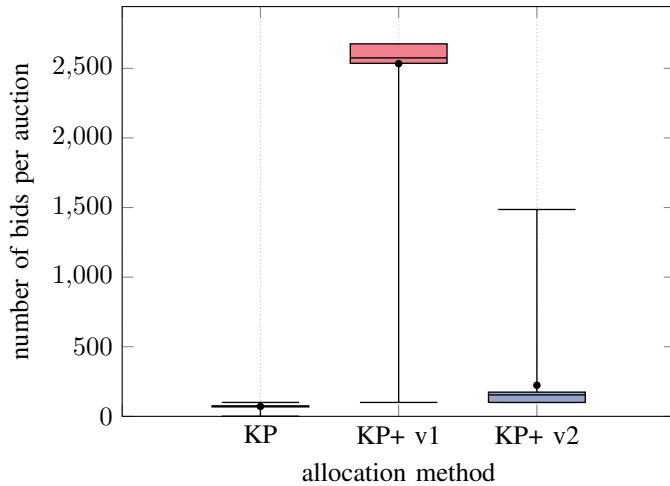


Fig. 3. Bid Requests for a large-scale Scenario of 100 Vehicles / 100 Tasks (5.000 Samples)

the architecture of a multi-agent system. With its ability to combine several simulation methods (*discrete-event*, *agent-based* and *system dynamics*), it fulfils, as the only software that features multi-method simulation modelling, the basic prerequisites for this research project. A combination of *discrete-event* and *agent-based* methods is proposed for the simulation study, as the performance of both methods can be examined in a risk-free environment.

The limitations of *AnyLogic* are of a general nature. As the software is proprietary and not open-source, the project depended on the provision by the *AnyLogic Company*. Furthermore, the project is bound to *Java* implementations for database applications or extensions (e.g. algorithms or model implementations). It is unlikely that either of these limitations will influence the feasibility of this project. General limitations that apply to the discrete-event simulation and agent-based modelling, relate to the level of detail with which the simulation is constructed (Mes et al. 2008; Heger and Voss 2018).

## V. DRAFT OF KARIS PRO+

Motivated by the informative simulation study of Hyland and Mahmassani (2018), our goal was to extend the Karis Pro task allocation method by permitting the reassignment of already allocated tasks. That is, vehicles that are currently driving to collect a load have to be able to diverge towards a new task and pass the one currently assigned to another vehicle to ensure that this task won't be late or forgotten. Also,

vehicles already carrying a load have to be able to change their reservation for the subsequent task if this is beneficial for the objective in the overall system.

The main difficulty is the decision when to reallocate an already assigned, but not initiated, task on to another vehicle, while allocating the new task to the formerly assigned vehicle so that the overall system benefits from this change. For this purpose, it is important to determine the opportunity costs, the costs accrued due to a reallocation, from an overall perspective. These costs in combination with the potential effort, more specifically time or distance a vehicle needs to travel, are used to formulate a key performance indicator to support the decision-making process.

### A. Opportunity Costs

Opportunity costs have to be considered whenever a vehicle has to pass an assigned task on to another vehicle. In our case, two potential outcomes may arise when a new task is auctioned and is subsequently assigned to one of the following vehicles:

- 1) A vehicle which is currently driving to pick up its assigned task (*pick up task*). This *pick up task* can be passed on to another vehicle and is then substituted by the new task.
- 2) A vehicle with a loaded task (*delivery task*) and a subsequent reservation (*reservation task*). Here, the *reservation task* can be passed on to another vehicle.

Due to the decentralised organisation, an increased flow of information is required between the vehicles in order to determine the opportunity costs: A vehicle, that intends to determine the opportunity costs of a currently assigned task, has to query all other vehicles for bids in an effort to determine which vehicle is going to perform the task. Due to the nature of this process, the process chain may well end in an infinity loop. That is, in order to ascertain these costs, the query is going to include vehicles that also have to calculate opportunity costs and therefore also have to query other vehicles. At this point, it is necessary to include only those vehicles in the calculation of opportunity costs that can provide an answer without again considering opportunity costs, i.e. vehicles in an idle state or a delivery state without reservations. Figure 2 shows the resulting bids of Version 1 of Karis Pro+ in red. The method can therefore be classified as not scalable (Lirkov and Margenov 2018). More than 1,600 bid calculations per tasks are required in a system with 80 vehicles in the system, in some cases even more.

Consequent to the primary feasibility assessment of Version 1 of Karis Pro+ described above, KP+ was adapted

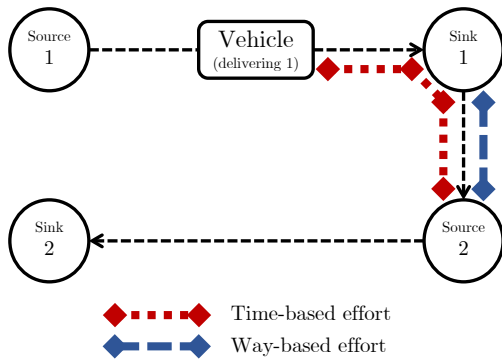


Fig. 4. Illustration of the time-based and the way-based Calculation of Effort

to Version 2 in order to reduce the number of necessary bid calculations. In Version 2 the opportunity costs are only determined if the best bidder has to determine the opportunity costs to establish the total costs for the system, i.e. *pickup* mode or *delivery* mode with *no reservation*. The reduction of the number of bid calculations from and KP+ v1 to KP+ v2 are substantial as is illustrated in Figure 2 and Figure 3.

### B. Priorities and Calculation of Effort

In the following simulation study, which was performed as a case study in the automotive sector, schedules play an important role. To ensure that all tasks are delivered on time, while utilising the temporal flexibility of each task in the planning process, Karis Pro+ operates with two types of priorities (*priority* and *no priority*). Whenever a task is in danger of being delayed, it will receive a priority flag, which will position it on the first ranks of the task list.

Furthermore, the calculation of the effort required for a vehicle to fulfil a specific task depends on its priority status. That is, all calculations for priority tasks are time-based, whereas the effort for no priority tasks is based on the distance a vehicle has to travel. This important differentiation ensures that priority tasks will be performed as fast as possible regardless of the distance it would need to travel. Both calculation methods are depicted in Figure 4. The time-based effort is measured in seconds from the present to the estimated time of arrival at the new task source (*Source 2* of Figure 4). However, the way-based effort is measured in meters from the current task's sink (*Sink 1*) to the new task's source (*Source 2*). Both calculation methods ensure that tasks *without priority* are delivered with the least traffic possible and tasks *with priority* as fast as possible.

Depending on this priority setting, each vehicle calculates its own effort in seconds or meters and – if requested – the opportunity costs. Table II illustrates the bid composition for different priority combinations for the newly auctioned task and the currently assigned task of the vehicle.

## VI. SIMULATION STUDY

The following section describes the simulation study, which was carried out to compare the performance of KP and KP+. The software *AnyLogic* version 8.6, was used to model the environment and all processes. In the upcoming subsections,

TABLE II. COMPOSITION OF BIDS REGARDING NEW TASKS AND CURRENT TASK'S PRIORITY

Priority		Bid Composition
New Mission (auction)	Current Mission (pickup / delivery with reservation)	
✓	✗	Only own Effort (time-based)
✗	✗	Own Effort + Opportunity Costs (both way-based)
✓	✓	Own Effort + Opportunity Costs (both time-based)
✗	✓	No bid calculated

the simulation setup, the parameters and the results will be presented in detail.

### A. Setup and Parameters

Tasks were carried out by a predetermined number of vehicles. All vehicles were positioned at an initial location at a random point beside the road network, also referred to as fit-point. Tasks appeared at random fit-points, however parameters were set to avoid clustering of such. Each task was assigned to a sink at a random free fit-point with a minimum linear distance of  $\frac{1}{3}$  of the total layout dimensions. The total number of tasks in the system was also limited by the parameter *Task Pool* for each simulation run. Therefore, whenever a task was completed, a new task appeared at another random fit-point. This logic ensured a constant utilisation of the vehicles, adjustable by the ratio between two parameters: *number of vehicles* and *size task pool*. When a task appeared, a random deadline between 20 and 30 minutes was set. This parameter was chosen to have a meaningful pressure on the system. Seven minutes prior to this deadline, a task was prioritised if had not already been collected by a vehicle. Priority tasks were collected by the closest vehicle and not, as is the case for non-prioritised tasks, by the vehicle with the least way-based effort. This parameter was chosen through preliminary studies in the same layout and ensures that tasks are not delayed.

### B. Scenarios

The study compares the Karis Pro and Karis Pro+ methods in scenarios incorporating a variety of scale and utilisation levels. We selected scenarios with 5, 30, 80 and 100 vehicles, thus providing differing scale levels and combined these with differently sized task pools to generate varying load factors. Three alternative task pools were chosen and applied to each vehicle pool size: one with the same number of tasks as number of vehicles, one with 20% fewer and one with 20% more tasks than vehicles. Each described scenario was calculated with both methods KP and KP+ and 5,000 tasks.

### C. Results and Discussion

The results of the simulation study are presented in the following section. The statistics are utilised to compare the task allocation methods Karis Pro and Karis Pro+ for the differently scaled scenarios.

Figure 5 illustrates the average utilisation for both allocation methods and the three different workload ratios. A vehicle is said to be utilised when it is either collecting a load or delivering a load. For all scenarios, KP+ has a lower utilisation

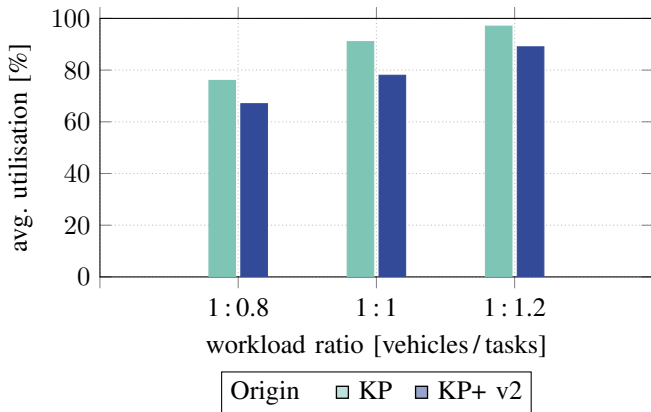


Fig. 5. Average Utilisation for each Workload and Allocation Method

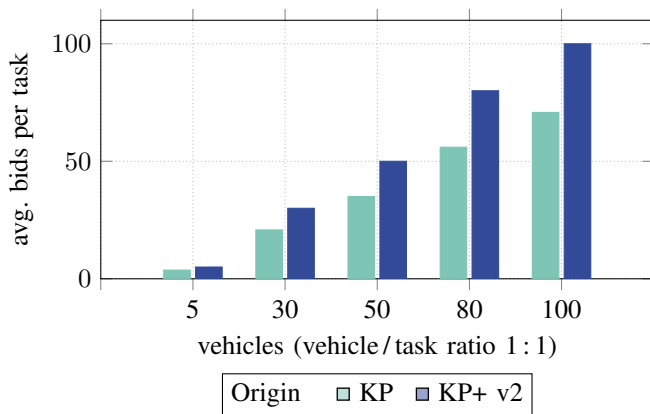


Fig. 6. Average Bids per Task for a Vehicle/Task Ratio of 1:1 (5,000 Samples)

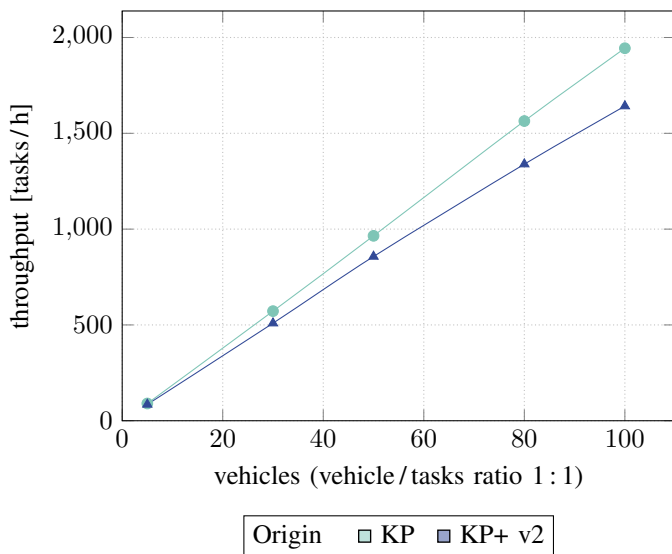


Fig. 7. Throughput measured in Tasks per Hour for KP and KP+ v2 (5,000 Samples)

than KP. No tasks were delayed, and all requirements were met in all of the scenarios. The lower utilisation rate can be explained by the fact that KP+ allocates tasks in advance. That is, vehicles already underway can be scheduled for a subsequent task, instead of activating an idle vehicle which would be associated with greater effort.

Corresponding with the lower utilisation of KP+, the throughput, given in tasks per hour, is also lower for KP+ than KP (see Figure 7). As all requirements are met in all KP+ scenarios, this finding can be deemed to be of secondary importance. KP+ makes use of the temporal flexibility of each task's deadline to execute it with the least necessary effort in terms of meters travelled per task. This finding is represented in Figure 8. The original KP method has a significant higher way-based effort to process tasks than KP+ (between 7% and 23%).

The required communication effort, represented by bid requests per task, was also evaluated. It might seem self-evident that KP+ should require more communication between the vehicles than KP, since vehicles bid in every mode and not only in *idle* or *delivery* mode. For a vehicle/task ratio of 1:1, the communication effort is illustrated in Figure 6.

## VII. CONCLUSION

Motivated by the contribution from Hyland and Mahmasani (2018), this study has introduced an alternative decentralised task allocation method based on Karis Pro (Colling et al. 2016), which has been proven to be superior, namely Karis Pro+. When auctioning a task, this version also considers vehicles that are either already performing a task or are reserved as a potential resource for a subsequent one. The aim of this study was to determine opportunity costs for the overall system and use them to make better decisions for reaching the objective. By using the same bidding system as Karis Pro, every vehicle is able to calculate the effort required for passing a task on to another vehicle. This feature improves the flexibility of the system by a substantial margin. The presented simulation

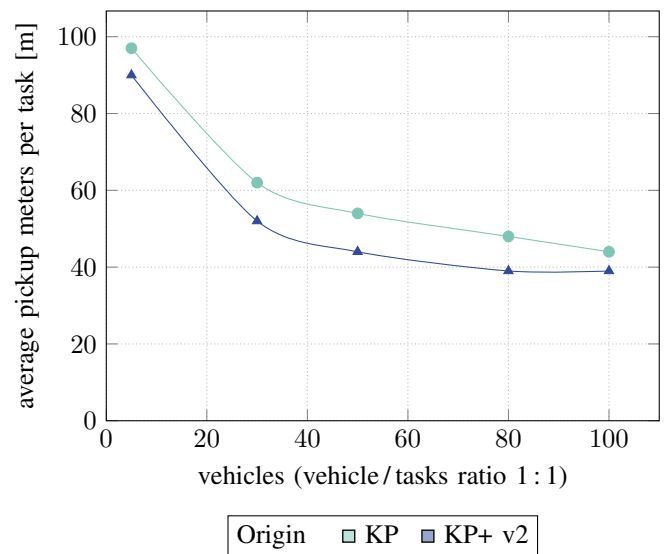


Fig. 8. Average Number of driven Pickup Meters per Task for a Vehicle/Task Ratio of 1:1 (5,000 Samples)

study supports the notion that using Karis Pro+ instead of Karis Pro leads to less empty drives and subsequently to a higher efficiency of the vehicle fleet. Consequently, it is suggested that a larger capacity buffer, resulting from this higher efficiency, may lead to higher flexibility for task peaks. This in turn may prevent traffic issues in large scale applications as the same goals may be achieved with lower traffic density, in comparison to the original allocation method of Karis Pro. Accordingly, the communication effort to reach this outcome is much higher than for the original method. Nevertheless, the developed simulation study demonstrated that Karis Pro+ in its final version can be considered to be a scalable method.

## REFERENCES

- Akerlof, George A. (1970). “The Market for Lemons: Quality Uncertainty and the Market Mechanism”. In: *The Quarterly Journal of Economics*, pp. 488–500.
- Bădică, Amelia, Costin Bădică, Florin Leon, and Ion Buligiu (2018). “Modeling and Optimization of Pickup and Delivery Problem Using Constraint Logic Programming”. In: *Large-scale scientific computing*. Ed. by Ivan D. Lirkov and Svetozar Margenov. Lecture Notes in Computer Science. Cham: Springer.
- Bellingham, John, Michael Tillerson, Arthur Richards, and Jonathan P. How (2003). “Multi-task Allocation and path Planning for Cooperating UAVs”. In: *Cooperative Control: Models, Applications and Algorithms*. Ed. by Sergiy Butenko, Robert Murphey, and Panos M. Pardalos. Cooperative Systems. Springer.
- Colling, Dominik, Sascha Ibrahimovic, Andreas Trenkle, and Kai Furmans (2016). “Dezentrale Auftragserteilung und -vergabe für FTF”. In: *Logistics Journal*.
- Dorri, Ali, Salil S. Kanhere, and Raja Jurdak (2018). “Multi-Agent Systems: A Survey”. In: *IEEE Access* 6, pp. 28573–28593.
- Gerkey, Brian P. and Maja J. Matarić (2002). “Sold!: auction methods for multirobot coordination”. In: *IEEE Transactions on Robotics and Automation* 18.5, pp. 758–768.
- (2004). “A formal analysis and taxonomy of task allocation in multi-robot systems”. In: *The International Journal of Robotics Research*.
- Heap, Bradford and Maurice Pagnucco (2013). “Repeated Sequential Single-Cluster Auctions with Dynamic Tasks for Multi-Robot Task Allocation with Pickup and Delivery”. In: *Multiagent System Technologies*. Ed. by Matthias Klusch. Lecture Notes in Computer Science. Berlin/Heidelberg: Springer.
- Heger, Jens and Thomas Voss (2018). “Optimal Scheduling of AGVs in a Reentrant Blocking Job-shop”. In: *Procedia CIRP* 67, pp. 41–45.
- Holt, Charles A. (1979). “Uncertainty and the Bidding for Incentive Contracts”. In: *American Economic Review*, pp. 697–705.
- Hyland, Michael and Hani S. Mahmassani (2018). “Dynamic autonomous vehicle fleet operations: Optimization-based strategies to assign AVs to immediate traveler demand requests”. In: *Transportation Research Part C: Emerging Technologies* 92, pp. 278–297.
- Innocenti, M., L. Pollini, and A. Bracci (2010). “Cooperative path planning and task assignment for unmanned air vehicles”. In: *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, pp. 121–131.
- Jazdi, N. (2014). “Cyber physical systems in the context of Industry 4.0”. In: *2014 IEEE International Conference on Automation, Quality and Testing, Robotics*. IEEE, pp. 1–4.
- Jennings, N. R., P. Faratin, A. R. Lomuscio, S. Parsons, Michael J. Wooldridge, and C. Sierra (2001). “Automated Negotiation: Prospects, Methods and Challenges”. In: *Group Decision and Negotiation*, pp. 199–215.
- Khamis, Alaa and Asser ElGindy (2012). “Minefield Mapping Using Cooperative Multirobot Systems”. In: *Journal of Robotics*.
- Khamis, Alaa, Ahmed Hussein, and Ahmed Elmogy (2015). “Multi-robot Task Allocation: A Review of the State-of-the-Art”. In: *Cooperative Robots and Sensor Networks 2015*. Ed. by Anis Koubâa and J. Ramiro Martínez-de Dios. Vol. 604. Studies in Computational Intelligence. Cham: Springer, pp. 31–51.
- Lagoudakis, Michail G., Marc Berhault, Sven Koenig, Pinar Keskinocak, and Anton J. Kleywegt (2004). “Simple auctions with performance guarantees for multi-robot task allocation”. In: *Intelligent Robots and Systems*.
- Leupold, Andreas, Falk Senger, Uwe Weber, and Tim Kappler (2018). “Industrial IoT”. In: *Digitale Welt* 2.3, pp. 78–84.
- Lirkov, Ivan D. and Svetozar Margenov, eds. (2018). *Large-scale scientific computing: 11th International Conference*. Lecture Notes in Computer Science. Cham: Springer.
- Liu, Chun and Andreas Kroll (2012). “A Centralized Multi-Robot Task Allocation for Industrial Plant Inspection by Using A\* and Genetic Algorithms”. In: *Artificial Intelligence and Soft Computing*. Ed. by Leszek Rutkowski, Marcin Korytkowski, Rafał Scherer, Ryszard Tadeusiewicz, Lotfi A. Zadeh, and Jacek M. Zurada. Berlin, Heidelberg: Springer, pp. 466–474.
- McIntire, Mitchell, Ernesto Nunes, and Maria L. Gini (2016). “Iterated Multi-Robot Auctions for Precedence-Constrained Task Scheduling”. In: *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. Ed. by Catholijn M. Jonker and Stacy Marsella.
- Mes, Martijn, Matthieu van der Heijden, and Jos van Hillegerberg (2008). “Design choices for agent-based control of AGVs in the dough making process”. In: *Decision Support Systems* 44.4, pp. 983–999.
- Moon, Sangwoo, David Hyunchul Shim, and Eunmi Oh (2015). “Cooperative Task Assignment and Path Planning for Multiple UAVs”. In: *Handbook of Unmanned Aerial Vehicles*. Ed. by Kimon P. Valavanis and George J. Vachtsevanos. Dordrecht: Springer, pp. 1547–1576.
- Nagarajan, Thareswari and Asokan Thondiyath (2013). “Heuristic based Task Allocation Algorithm for Multiple Robots Using Agents”. In: *Procedia Engineering* 64, pp. 844–853.
- Nunes, Ernesto and Maria L. Gini (2015). “Multi-Robot Auctions for Allocation of Tasks with Temporal Constraints”. In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- Nunes, Ernesto, Marie Manner, Hakim Mitiche, and Maria L. Gini (2017). “A Taxonomy for Task Allocation Problems

- with Temporal and Ordering Constraints”. In: *Robotics and Autonomous Systems* 90, pp. 55–70.
- Schwarz, Christoph (2014). “Untersuchung zur Steigerbarkeit von Flexibilität, Performanz und Erweiterbarkeit von Fahrerlosen Transportsystemen durch den Einsatz dezentraler Steuerungstechniken”. Dissertation. Oldenburg: Carl-von-Ossietzky-Universität Oldenburg.
- Schwarz, Christoph, Jurij Schachmanow, Jürgen Sauer, Ludger Overmeyer, and Georg Ullmann (2013). “Selbstgesteuerte Fahrerlose Transportsysteme”. In: *Logistics Journal*.
- Tovey, Craig, Michail G. Lagoudakis, Sonal Jain, and Sven Koenig (2005). “The generation of bidding rules for auction-based robot coordination”. In: *Multi-Robot Systems. From Swarms to Intelligent Automata Volume III*. Ed. by Lynne E. Parker, Frank E. Schneider, and Alan C. Schultz.
- Trenkle, Andreas (2016). “Karis PRO – Autonomer Materialtransport für flexible Intralogistik: Abschlussbericht des BMBF-Verbundforschungsprojektes”. In:
- Versteegt, C. and A. Verbraeck (2002). “Holon control of large-scale automated logistic systems”. In: *Proceedings of the IEEE 5th International Conference on Intelligent Transportation Systems*. IEEE, pp. 898–903.
- Windelinckx, Neil and Malcolm Strens (2004). “Dynamic scheduling of multiple agents for a heterogeneous task stream”. In: *Proceedings of the AISB 2004*.

#### AUTHOR BIOGRAPHIES

**Maximilian Selmair** is a doctoral student at the University of Plymouth. Recently employed at the SimPlan AG, he was in charge of projects in the area of material flow simulation. He is currently working on his doctoral thesis as a part of a fellowship of the BMW Group. His email address is: maximilian.selmair@bmw.de and his website can be found at maximilian.selmair.de.

**Prof. Dr. Klaus-Jürgen Meier** holds the professorship for production planning and logistic systems in the Department of Engineering and Management at the University of Applied Sciences, Munich, and he is the head of the Institute for Production Management and Logistics (IPL). His email address is: klaus-juergen.meier@hm.edu.

# OPTIMAL PRODUCTION VOLUME OF RUBBER GLOVES MOLD FOR RUBBER GLOVES PRODUCTION PLANNING

Tuanjai Somboonwivat  
Chorkaew Jaturanonda  
Nattapong Chotpan  
Department of Production Engineering  
King Mongkut's University of  
Technology Thonburi (KMUTT)  
Thailand  
E-mail: tuanjai.som@kmutt.ac.th

Kanogkan Leerojanaprapa  
Statistics Department  
King Mongkut's Institute of  
Technology Ladkrabang (KMITL)  
Thailand  
E-mail: klkanogk@kmitl.ac.th

## KEYWORDS

Production planning, Rubber glove, Rubber glove Mold.

## ABSTRACT

The production process of rubber glove is the continuous production line which the hand-shaped ceramic molds are installed and dipped into the concentrated latex to form the rubber gloves. Since, the rubber gloves are diverse in terms of size, surfaces and types of latex, the multiple molds and latex are changed and used in a production process in order to produce rubber gloves from customer requirement. Thus, the changing of mold is extremely complex for production ratio which the rubber glove production planning will be employed to meet anticipate customer orders. This research therefore develops the mathematical model to find the optimal quantity of molds and rubber gloves production planning. Finally, the developed model is applied to an example data set to find the minimum total volume of rubber gloves in every mold for keeping the minimal volume of inventory.

## INTRODUCTION

Rubber gloves industry has become increasing significant for Thai economy. At present, a large volume of natural latex is used for producing rubber gloves which extracted from rubber trees. Rubber is one of the important industrial crops for Thailand, both for the local consumption and the global market. From the export data of Thai Custom Department in 2014 it was found that rubber gloves had contributed an enormous export values to Thai economy. The uses of gloves as part of infection control within healthcare will be increased extensively in the near future, as at present people are more and more aware of their health and cleanliness. The growth of sale will also affect the manufactures to serve their customer needs.

Rubber gloves industry has a complicate and complex process in production planning due to several of type and size of products. Moreover, each type of glove uses different raw materials to produce. The process of producing rubber gloves is forming by dipping each size

hand-shaped ceramic mold into a tank full of concentrated latex with the length of time. Moreover, the machines will be run all the time. The volume of production in each batch will depend on a number of hand-shaped ceramic molds which are set up within each machine by size and type of mold. Therefore, in production planning process, it is obligatory to find a number of hand-shaped ceramic molds which are fit with orders from customers. Excess or shortage items should be minimal in each type and size of production run. If the production planning process does not match with customers' orders, it is necessary to stop the production run in order to set up the machine again by changing molds. This will result in losing time and budget in setting up the machines.

Most of the literature in the area of production management for the rubber glove industry is related to the structure of the rubber industry (Haan et al. 2003), production improvement (Jirasukprasert et al. 2012) and economics and environment (Rattanapana et al. 2012). The researches on production planning for the rubber gloves manufacturing are still limited. Klomsae et al. 2012 present an applied mathematical model for rubber latex industry decision planning, that covers purchase and storage over a multi-period timeframe, with due consideration of product aging and deterioration through each time period. However, there has not been research related to the production planning of rubber glove products involving the mold. Thus, this paper presents the production planning and scheduling for rubber gloves with mixed mold sizes and types.

Most production planning and scheduling has been done by applying mathematical models (Hsu et al. 2011, Wen-Chiung et al. 2012 and Yan et al. 2013). Birger R. et. al. 2013 present the mathematical model for multi-product multi-period aggregate production-distribution planning problem with mould sharing between plants. This paper aims to create a mathematical model for rubber glove production planning and scheduling for multi-glove products, multi-molds, multi-concentrated latex and multi-production lines for multi-periods to plan production in order to meet the customer requirement.

The paper is structured as follows: In section 2, we provide general information about type of rubber gloves and rubber gloves production machine describe the production planning process. Then, the problem statement is discussed in section 3 and the mathematical model of the problem is developed in section 4. A numerical example and its computational results are presented in Section 5. Finally, concluding remarks and further work are provided in the last section.

## RUBBER GLOVES INFORMATION

The types of rubber gloves, hand-shaped ceramic molds and concentrated latex are described in this section.

### Types of Rubber Gloves

Generally, here are three main types of rubber glove i.e. 1) medical glove, 2) industrial glove and 3) household glove. Details are as follows:

#### a) Medical Glove

Medical gloves use during medical procedures to prevent contamination between caregivers and patients. Medical glove can be divided into two major types which are surgical and examination. Surgical gloves are generally sterile and feature extra long reinforced cuffs. Generally, these gloves are thicker for use in higher-risk clinical applications for extra protection. So, high technology process in production and sensitive procedures are necessary to make sure that manufacturers of these devices have a higher standard. Examination gloves are available as either sterile or non-sterile which made from 100% synthetic latex, both powder-free and powdered. Medical gloves are regulated by the Food and Drug Administration (FDA) to make sure that manufacturers can meet performance criteria such as leak and tear resistance.

#### b) Industrial Glove

There are various types of industrial gloves which are produced to protect against a wide variety of hazards. Normally, determining which type of glove to use is depended upon the duration of the job, the type of conditions or the environment (wet or dry). These gloves are designed to ensure employee safety and sanitary conditions in the workplace and provide both strength and chemical protection. For example, heat or cut resistant gloves have different properties and compositions from chemical and oil resistant glove.

#### c) Household Glove

Household glove are generally used for domestic cleaning and food processing purposes. It provides ergonomic design to ensures excellent abrasion as well as cut and tear resistance. There are different designs and colour of this type of glove in order to attract housewives. The best rubber gloves should be durable, whilst allowing users to accomplish their tasks effectively.

## Hand-shaped Ceramic Molds

A number of gloves to be produced depend on a number of molds which are set up in a machine. Normally, there are many kinds of rubber glove, for instance, just smooth skin or "dot" on the fingertips or palms that let users can touch device without smudging or scratching it. Moreover, there are many sizes of rubber gloves which suit for users' hand such as XS, S, M, L, XL or XXL. In each mold conveyor, it can be mixed every size and type of rubber gloves in one production line. Hence, planners should find out the exactly molds in their factories before making a production scheduling. Table 1 shows an example of molds for each item in a case study of this research.

Table 1: An Example of Molds for Each Item

Number of hand-shaped ceramic molds						
Type\size	XS	S	M	L	XL	XXL
Smooth Skin	10,000	20,000	50,000	50,000	20,000	10,000
Dots Balm	10,000	10,000	20,000	20,000	10,000	10,000
Dots Finger	10,000	10,000	10,000	10,000	10,000	10,000

### Types of Concentrated Latex

Natural rubber is obtained from different species of rubber trees in form of latex which is called normal or field latex. The latex is then placed into a centrifuge to remove some of water and increase rubber content of the latex. After centrifuging, the material is known as concentrated latex, which contains about 60% rubber, are then compounded by adding other raw material ingredients to achieve the desired performance characteristics. On the other hand, synthetic latex is produced from a petrochemical which is developed as an alternative to natural latex for some people who are allergic reaction to natural latex rubber. The 4 types of latex used in this study are Natural latex, Synthetic white latex, Synthetic blue latex, and Synthetic black latex.

## PROBLEM STATEMENT

The productions planning process of rubber gloves starts from the marketing department receives the orders from customers and sends them to the production planning department for scheduling. Normally, at the first step a planner will rearrange schedule according to due date and inform a plan to other departments such as logistics and transportation department. Afterwards, the planner will check on-hand inventory and calculate actual volumes that are needed to produce. It is necessary to consider capacity of each size in all production lines in real time in order to create production schedule of each machine. If the machine is not processing the same type of rubber gloves, it will be terminated to change ceramic molds; otherwise, that orders are cancelled. After the master production

schedule (MPS) is set up, it will be returned to customers to confirm delivery time.

At present, a planner can get anticipated demand in each type and size of product in advance. A number of molds for each type and size to be fixed with each machine as well as running time process are calculated to be ensured that the products will be sent to customers on time. The amount of products depends on a number of molds which are set up at each machine by taking size and type of molds into consideration. Such process is very complicated and takes a long time for the production planner in operating and sometimes this can cause a mistake. Such optimization problem can effectively be solved using mathematical modeling. Figure 1 presents the conveyors of production lines which the molds are installed.

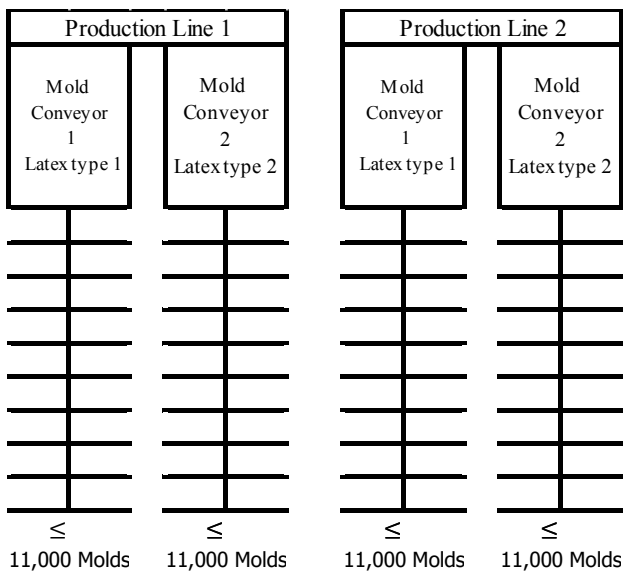


Figure 1: Mold Conveyors of Production Lines

Therefore, the production planning of rubber gloves is a process of bringing anticipated demand one month ahead to find the production volume in order to set up machines. Each machine is required to produce rubber gloves upon the customers demand. The followings are the conditions and limitation of rubber glove production machine.

- The number of mold for each machine is limited.
- Each conveyor has the same piece of mold and run simultaneously.
- Each type and size of molds is limited.
- Each machine can mix 2 types of latex. In mold conveyor 1 and mold conveyor 2 must be the same latex.

Table 2 summarizes the conditions.

Table 2: Conditions of Rubber Glove Production Machine

Production line 1	
The machine runs 24 hours per day for 30 days for each production period.	
Rubber glove mold conveyor 1	Rubber glove mold conveyor 2
<ul style="list-style-type: none"> <li>- It can be used only one type of latex.</li> <li>- The maximum hand-shaped ceramic molds are less than 11,000 pieces. Therefore, the maximum capacity per day is less than <math>11,000 \times 26.81</math> (production rate of one ceramic mold per day) or 294,910 pieces per day.</li> <li>- Any size of ceramic molds can be mixed, however; ceramic molds are limited.</li> </ul>	<ul style="list-style-type: none"> <li>- It can be used only one type of latex.</li> <li>- The maximum hand-shaped ceramic molds are less than 11,000 pieces. Therefore, the maximum capacity per day is less than <math>11,000 \times 26.81</math> (production rate of one ceramic mold per day) or 294,910 pieces per day.</li> <li>- Any size of ceramic molds can be mixed, however; ceramic molds are limited.</li> </ul>

## MATHEMATICAL MODEL

Details of indices, decision variables, parameters, objective functions, and constraints in production planning for rubber gloves in one month planning horizon are as follows:

### Indices

- $i$  Type and size of hand-shaped ceramic mold;  
 $i = 1, 2, 3, \dots, I$
- $j$  Product type of latex;  
 $j = 1, 2, 3, \dots, J$
- $k$  Product type mold conveyor;  
 $k = 1, 2, 3, \dots, K$
- $l$  Production line;  
 $l = 1, 2, 3, \dots, L$
- $t$  Date;  
 $t = 1, 2, 3, \dots, 30$

### Parameters

- $d_{ijt}$  Demand of rubber gloves for molds ' $i$ ' using latex ' $j$ ' in date ' $t$ '
- $s_{ijt}$  Inventory of rubber gloves for molds ' $i$ ' using latex ' $j$ ' in date ' $t$ '
- $c_{kl}$  Maximum capacity per day for mold conveyor ' $k$ ' in production line ' $l$ '
- $m_i$  Maximum volume of rubber gloves using molds ' $i$ '

### Decision Variables

- $x_{ijklt}$  Quantity/Volume of rubber gloves from mold ' $i$ ' using latex ' $j$ ' by typing of mold conveyor ' $k$ ' in production line ' $l$ ' in date ' $t$ '
- $y_{jklt}$  Decision variable which will take on only possible states i.e. yes (1) or no (0) production by using latex ' $j$ ', typing of mold conveyor ' $k$ ' in production line ' $l$ ' in date ' $t$ '



## Objective function

The objective function of the problem is to find the minimum total volume of rubber gloves in every mold 'i' and latex 'j' in 30 days, as shown in equation (1).

$$\text{Min } Z = \sum_{i=1}^I \sum_{j=1}^J s_{ij30} \quad (1)$$

where

$$s_{ijt} = x_{ijklt} \cdot y_{jkl} + s_{ijt-1} - d_{ijt} \quad ; \forall i, j, k, l, t \quad (2)$$

## Constraints

1. Quantity/Volume of rubber gloves from mold 'i' using latex concentrate 'j' by typing of mold conveyor 'k' in production line 'l' in date 't' is less than or equal to maximum capacity per day for mold conveyor 'k' in production line 'l'.

$$\sum_{i=1}^I \sum_{j=1}^J x_{ijklt} \leq c_{kl} \quad ; \forall k, l, t \quad (3)$$

2. Quantity/Volume of rubber gloves from mold 'i' using latex concentrate 'j' by typing of conveyor 'k' in production line 'l' in date 't' is less than or equal to maximum volume of rubber gloves from using molds 'i'.

$$\sum_{j=1}^J \sum_{k=1}^K \sum_{l=1}^L x_{ijklt} \leq m_i \quad ; \forall i, t \quad (4)$$

3. The production of rubber gloves in every mold 'i' in each conveyor belt 'k' in production line 'l' must be the same latex concentrate 'j' only.

$$\sum_{j=1}^J y_{jkl} = 1 \quad ; \forall k, l, t \quad (5)$$

4. Shortage of rubber gloves for molds 'i' using latex 'j' in date 't' is not allowed.

$$s_{ijt} \geq 0 \quad ; \forall i, j, t \quad (6)$$

5. Constraints describing the decision variables.

$$x_{ijklt} \geq 0 \quad ; \forall i, j, k, l, t \quad (7)$$

$$y_{jkl} \in \{0, 1\} \quad ; \forall j, k, l, t \quad (8)$$

The results from the production planning in each month using equation (1) to (8) will keep minimal on the volume of inventory of rubber gloves using mold 'i' in every latex 'j' in the final date of the time horizon. The schematic diagram in Table 3 shows the production quantity and demand of rubber gloves.

## NUMERICAL EXAMPLE

This section demonstrates a numerical example by applying mathematical model described in section 4 using a case company in Thailand. The objective of the study is to find the number of rubber gloves volume to

be produced which keep inventory in the system at the minimal by considering the requirement of customers in each item and due date. An example of customer requirement of the rubber gloves for the smooth skin, XS size and natural latex gloves (Table 4). The first order is due on the fifteen of the month and the total requirement is 5,000,000 pieces.

Table 4: Rubber Glove Production Demand

Customer Demand				
Type	Size	Type of Concentrated Latex	Volume (piece)	Due Date
Smooth Skin	XS	natural latex	5,000,000	15
Smooth Skin	S	natural latex	5,000,000	15
Smooth Skin	M	natural latex	5,000,000	15
Smooth Skin	L	natural latex	5,000,000	15
Smooth Skin	XL	natural latex	5,000,000	15
Smooth Skin	XXL	natural latex	5,000,000	15
Smooth Skin	XS	natural latex	2,000,000	20
Dots Balm	L	natural latex	5,000,000	20
Dots Balm	XL	natural latex	5,000,000	20

The results of production planning and scheduling of rubber gloves will be displayed in terms of decision variables of the volume of gloves to be produced using mold 'i' latex 'j' for mold conveyor 'k' on production line 'l' at date 't'. An example of on hand inventory of rubber gloves is zero inventory in first period. All data will be computed under production ratio of a machine using Excel solver. The results of the volume of molds to be set up for a mold conveyor 'k' on production line 'l' at date 't' are demonstrated in Table 5. By applying these formula in a production planning procedure and adjusting mold by following computational results would result in the total of inventory volumes for every mold and latex in 30 days at the minimum.

## CONCLUSION AND FURTHER STUDY

This paper, the mathematical model was formulated to minimize total volume of rubber gloves under limited resource constraints. A numerical example of case study was presented. The short-term scheduling time horizon for rubber gloves production is one month, however; this model can be extended for multi-stage decision making in the future. The proposed production planning and scheduling model can improve and organize several management decisions for producing, carrying inventories, and balancing utilization of machines in the short-term planning time horizon. However, for analysis process, it may need person who can understand mathematical precisely and take time to solve such problem. Consequently, developing decision support system (DSS) can assist and support decision makers in accommodating changes in the input information.

Table 3: The Production Volume Addressed in Mathematical Model for Rubber Glove Production Planning

Rubber Gloves Production Volume of Mold $i$ , Latex $j$ , Mold Conveyer $k$ , and Production Line $l$ , in Date $t$ (piece) $X_{ijklt}$								Rubber Gloves Demand of Mold $i$ , Latex $j$ , in Date $t$ (piece) $D_{ijt}$				
Mold	Latex	Mold Conveyer	Production Line	Date of Production				Date of Rubber Glove Demant				
				$t$				$t$				
$i$	$j$	$k$	$l$	1	2	...	30	1	2	...	30	
$1$	$1$	$1$	1	$X_{11111}$	$X_{11112}$	...	$X_{111130}$	$D_{111}$	$D_{112}$	...	$D_{1130}$	
			2	$X_{11121}$	$X_{11122}$	...	$X_{111230}$					
			:	...	...	...	...					
			$l$	$X_{111l1}$	$X_{111l2}$	...	$X_{111l30}$					
			$2$	1	$X_{11211}$	$X_{11212}$	...					$X_{112130}$
				2	$X_{11221}$	$X_{11222}$	...					$X_{112230}$
		:		...	...	...	...					
		$l$		$X_{112l1}$	$X_{112l2}$	...	$X_{112l30}$					
		$k$		1	$X_{11k11}$	$X_{11k12}$	...					$X_{11k130}$
				2	$X_{11k21}$	$X_{11k22}$	...					$X_{11k230}$
			:	...	...	...	...					
			$l$	$X_{11kl1}$	$X_{11kl2}$	...	$X_{11kl30}$					
	$2$		$1$	1	$X_{12111}$	$X_{12112}$	...	$X_{121130}$	$D_{121}$	$D_{122}$	...	$D_{1230}$
				2	$X_{12121}$	$X_{12122}$	...	$X_{121230}$				
		:		...	...	...	...					
		$l$		$X_{121l1}$	$X_{121l2}$	...	$X_{121l30}$					
		$2$		1	$X_{12211}$	$X_{12212}$	...	$X_{122130}$				
				2	$X_{12221}$	$X_{12222}$	...	$X_{122230}$				
			:	...	...	...	...					
			$l$	$X_{122l1}$	$X_{122l2}$	...	$X_{122l30}$					
			$k$	1	$X_{12k11}$	$X_{12k12}$	...	$X_{12k130}$				
				2	$X_{12k21}$	$X_{12k22}$	...	$X_{12k230}$				
		:		...	...	...	...					
		$l$		$X_{12kl1}$	$X_{12kl2}$	...	$X_{12kl30}$					
:	:	:		:	:	:	:	:	:	:		
$j$	$k$	$l$		$X_{1jkl1}$	$X_{1jkl2}$	...	$X_{1jkl30}$	$D_{1j1}$	$D_{1j2}$	...	$D_{1j30}$	
$2$	$1$	$1$	1	$X_{21111}$	$X_{21112}$	...	$X_{211130}$	$D_{211}$	$D_{212}$	...	$D_{2130}$	
			2	$X_{21121}$	$X_{21122}$	...	$X_{211230}$					
			:	...	...	...	...					
			$l$	$X_{211l1}$	$X_{211l2}$	...	$X_{211l30}$					
			$2$	1	$X_{21211}$	$X_{21212}$	...					$X_{212130}$
				2	$X_{21221}$	$X_{21222}$	...					$X_{212230}$
		:		...	...	...	...					
		$l$		$X_{212l1}$	$X_{212l2}$	...	$X_{212l30}$					
		$k$		1	$X_{21k11}$	$X_{21k12}$	...					$X_{21k130}$
				2	$X_{21k21}$	$X_{21k22}$	...					$X_{21k230}$
			:	...	...	...	...					
			$l$	$X_{21kl1}$	$X_{21kl2}$	...	$X_{21kl30}$					
	$2$		$1$	1	$X_{22111}$	$X_{22112}$	...	$X_{221130}$	$D_{221}$	$D_{222}$	...	$D_{2230}$
				2	$X_{22121}$	$X_{22122}$	...	$X_{221230}$				
		:		...	...	...	...					
		$l$		$X_{221l1}$	$X_{221l2}$	...	$X_{221l30}$					
		$2$		1	$X_{22211}$	$X_{22212}$	...	$X_{222130}$				
				2	$X_{22221}$	$X_{22222}$	...	$X_{222230}$				
			:	...	...	...	...					
			$l$	$X_{222l1}$	$X_{222l2}$	...	$X_{222l30}$					
			$k$	1	$X_{22k11}$	$X_{22k12}$	...	$X_{22k130}$				
				2	$X_{22k21}$	$X_{22k22}$	...	$X_{22k230}$				
		:		...	...	...	...					
		$l$		$X_{22kl1}$	$X_{22kl2}$	...	$X_{22kl30}$					
:	:	:		:	:	:	:	:	:	:		
$j$	$k$	$l$		$X_{2jkl1}$	$X_{2jkl2}$	...	$X_{2jkl30}$	$D_{2j1}$	$D_{2j2}$	...	$D_{2j30}$	
:	:	:	:	:	:	:	:	:	:	:		
$i$	$j$	$k$	$l$	$X_{ijkl1}$	$X_{ijkl2}$	...	$X_{ijkl30}$	$D_{ij1}$	$D_{ij2}$	...	$D_{ij30}$	

Table 5: Mold and Type of Concentrated Latex Installation Result

Mold and Latex Install Result (Piece)																						
Production Line	Rubber Glove Production Machine		Smooth Skm Mold					Dots Balm Mold					Dots Finger Mold									
	Mold Conveyer	Type of Concentrated Latex	XS	S	M	L	XL	XXL	XS	S	M	L	XL	XXL	XS	S	M	L	XL	XXL		
1	1	Natural latex	1,467	2,933	733		1,467	6,600					1,467									
		Synthetic white latex																				
		Synthetic blue latex																				
		Synthetic black latex																				
2	2	Natural latex	1,630	1,731	2,078	1,100	1,467	4,094				1,100	1,467									
		Synthetic white latex																				
		Synthetic blue latex																				
		Synthetic black latex																				
2	1	Natural latex	2,933	1,467	2,933	733	1,945	1,976				733	1,945									
		Synthetic white latex																				
		Synthetic blue latex																				
		Synthetic black latex																				
2	2	Natural latex	3,667	2,200	4,787		1,100	1,813					1,100									
		Synthetic white latex																				
		Synthetic blue latex																				
		Synthetic black latex																				
3	1	Natural latex	4,828	1,467	2,200	713	2,200	1,467				713	2,200									
		Synthetic white latex																				
		Synthetic blue latex																				
		Synthetic black latex																				
3	2	Natural latex	3,300	2,933	5,093	0	1,467	1,874						733								
		Synthetic white latex																				
		Synthetic blue latex																				
		Synthetic black latex																				

## REFERENCES

- Birger R.; D. Wout and A. El-Houssaine. 2013. "A Matheuristic for Aggregate Production–distribution Planning with Mould Sharing". *International Journal of Production Economics*, 145, Issue 1, 29-37.
- Dehau, X. and Y. Dar-Li. 2013. "Makespan Minization for Two Parallel Machines Model with a Periodic Availability Constraint : Mathematical Programming Model, Average-Case Analysis, and Anomalies". *International Journal of Applied Mathematical Modelling*, 7561-7567.
- Jirasukprasert, P.; J.A. Garza-Reyes; H. Soriano-Meier; H., and L. Rocha-Lona. 2012. "A Case Study of Defects Reduction in a Rubber Gloves Manufacturing Process by Applying Six Sigma Principles and DMAIC Problem Solving Methodology." *Proceedings of the 2012 International Conference on Industrial Engineering and Operations Management*, 472-481.
- Haan, J. de.; G. De. Groot; E. Loo and M. Ypenburg. 2003. "Flows of Goods or Supply Chains; Lessons from the Natural Rubber Industry in Kerala, India." *International Journal of Production Economics*, Vol. 81-82, 185-194.
- Hsu, C.J.; W.H. Kuo and D.L. Yang. 2011. "Unrelated Parallel Machine Scheduling with Past-sequence-dependent Setup Time and Learning". *Applied Mathematical Modelling* 35, 1492–1496.
- Klomsae, S.; T. Somboonwiwat and W. Atthirawong. 2012. "Optimal multi-period Planning for Purchase and Storage of Rubber Latex with Perishability Constraints." *Proceedings of the 7<sup>th</sup> International Congress on Logistics and SCM Systems*, 244-250.
- Rattanapana, C.; T.T. Suksaroj, W.Ounsanehab. 2012. "Development of Eco-efficiency Indicators for Rubber Glove Product by Material Flow Analysis". *Procedia - Social and Behavioral Sciences* 40, 99 – 106.
- Wen-Chiung, L.; C. Mei-Chi and Y. Wei-Chang. 2012. "Uniform Parallel-machine Scheduling to Minimize Makespan with Position-based Learning Curves". *International Journal of Information Computer and Industrial Engineering*, 813-818.
- Yan, H.S.; H.X. Wang and X.D. Zhang. 2013. "Simultaneous Batch Splitting and Scheduling on Identical Parallel Production Lines". *Information Science* 221, 501–519.

## AUTHOR BIOGRAPHIES



**TUANJAI SOMBOONWIWAT** is an associate Professor in the Industrial Management section, Department of Production Engineering, King Mongkut's University of Technology Thonburi, Thailand. She received her Ph.D. in Industrial Engineering from Corvallis, Oregon State University, USA. Her research interests include green supply chain and logistics, business process and applications of operations research. Her e-mail address is : [tuanjai.som@kmutt.ac.th](mailto:tuanjai.som@kmutt.ac.th).



**CHORKAEW JATURANONDA** is a faculty member at Department of Production Engineering, King Mongkut's University of Technology Thonburi, Thailand. She received a bachelor's degree in Applied Mathematics from King Mongkut's University of Technology Ladkrabang (Thailand), a master's degree in Industrial Engineering from University of Texas at Arlington (USA), and a Ph.D. degree in industrial engineering from Sirindhorn International Institute of Technology, Thammasat University (Thailand). Her research interests include applied operations research, logistics management, and Decision Support Systems. Her e-mail address is : [chorkaew.jat@kmutt.ac.th](mailto:chorkaew.jat@kmutt.ac.th)



**KANOGKAN LEEROJANAPRAPA** was born in Bangkok, Thailand and went to the Thammasat University, where she studied Applied Statistics and obtained her degree in 1999. She continued to study Master degree in Statistics in Chulalongkorn University and obtained her degree in 2002. She worked for four years for King Mongkut's Institute of Technology Ladkrabang (KMITL) before doing her PhD in 2008, University of Strathclyde, UK. After her graduation, she returned to KMITL where she is now a lecturer in Statistics department. Her e-mail address is : [klkanogk@kmitl.ac.th](mailto:klkanogk@kmitl.ac.th)



**NATTAPONG CHOTPAN** was a graduate student in Industrial and Manufacturing Systems Engineering program at the Department of Production Engineering, King Mongkut's University of Technology Thonburi, Thailand. He received his bachelor's degree in Industrial Engineering from Ramkhamhaeng University. His research interests are in supply chain management, production planning and optimization model. His e-mail address is : [nattapong-ch@hotmail.com](mailto:nattapong-ch@hotmail.com).

# A literature database on ecological sustainability in hierarchical production planning

Hajo Terbrack<sup>a\*1</sup>, Ingo Frank<sup>b2</sup>, Frank Herrmann<sup>b3</sup>, Thorsten Claus<sup>a4</sup>, Marco Trost<sup>a5</sup> and Matthias Götz<sup>a6</sup>

<sup>a</sup> *International Institute (IHI) Zittau, Technische Universität Dresden, Dresden, Germany.*

<sup>b</sup> *Innovation and Competence Centre for Production Logistics and Factory Planning, OTH Regensburg, Regensburg, Germany.*

\* *Corresponding author*

<sup>1</sup> *Hajo.Terbrack@mailbox.tu-dresden.de*, <sup>2</sup> *Ingo.Frank@extern.oth-regensburg.de*,

<sup>3</sup> *Frank.Herrmann@oth-regensburg.de*, <sup>4</sup> *Thorsten.Claus@tu-dresden.de*, <sup>5</sup> *Marco.Trost@tu-dresden.de* and

<sup>6</sup> *Matthias.Goetz@tu-dresden.de*.

## Keywords

Sustainability, ecology, production planning, literature search.

## Abstract

This article describes a systematic literature search of research articles on the topics of ecological sustainability and production planning. Over 900 research articles deal intensively with these topics. These articles were catalogued and included in a database. They can be evaluated with an analysis tool, developed by the research group. The tool is available free of charge via a website of the research group.

## 1. Introduction

Sustainable concerns are increasing in almost every part of human society. Mainly because of the consumption of resources such as water or energy, production companies should produce or manufacture respectively more sustainably. This might be improved by planning.

In this article we limit ourselves to ecological sustainability and hierarchical production planning.

Already a first research in renowned scientific journals like the "Journal of Cleaner Production" or the "International Journal of Production Economics" show that so many articles on ecological sustainability in hierarchical production planning have been published that they cannot be covered in a single scientific publication. Since we are working on several different research questions for planning in our research group, we have performed a literature search for relevant articles. We have catalogued these articles and stored them in our own database. In addition, we have developed an analysis tool with which articles can be selected for specific research questions.

The remainder of this article is structured as follows. Section 2 contains the scope of the literature database and section 3 the methodology for the literature search. Information regarding the database and the analysis tool are given in section 4 and the article is summarized in section 5.

## 2. Scope of the literature database

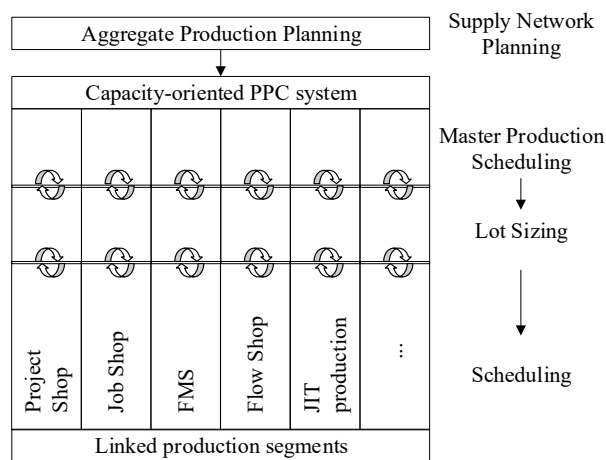
Drivers for sustainable manufacturing are, for example, stakeholders such as employees, environmental activists and government bodies. Other factors such as limited availability of resources and rising energy costs lead to a greater significance of sustainability for companies. The goal to improve sustainability was already published in 1987 by the World Commission on Environment and Development, known as Brundtland Report (see Brundtland et al. 1987). Further, the U.S. Environmental Protection Agency define sustainable manufacturing as "[...] the creation of manufactured products through economically-sound processes that minimize negative environmental impacts while conserving energy and natural resources. Sustainable manufacturing also enhances employee, community and product safety." (U.S. Environmental Protection Agency 2020). Thus, it can be concluded that sustainable production should include at least two elements of the "triple bottom line", i.e. the economic pillar together with the environmental or social pillar (Akbar and Irohara 2018).

In present, the influence of industrial production on the environment is huge. In 2018, the industry sector accounted for 37% of global final energy usage and for 24% of global emission (IEA 2020). Furthermore, in Germany in 2018, 15% of the total waste has incurred in industry (Destatis 2020). In 2013, the industrial sector was responsible for more than 14% of the total water use in Germany (Eurostat 2017). In this respect, the research on sustainable manufacturing address the improvement of products, machines and processes.

According to Akbar and Irohara 2018 production planning and control (PPC) offers considerable potential for sustainable improvements, because a more effective and efficient use of resources is achieved. This is usually achieved without new investments and can be implemented immediately. A majority of the sustainable research for PPC focus on the ecological dimension – examples are: Absi et al. 2013, Akbar and Irohara 2018, Battini et al. 2014, Bouchery et al. 2012. With an adapted PPC, production programs and resource allocations can be used to directly improve related aspects such as energy

consumption, waste production or emission. In the following, the planning hierarchy and ecological sustainability are described more in detail, since their contents are relevant for literature research and cataloguing within the database.

The operational PPC, as implemented in commercially available Enterprise Resource Planning (ERP) systems, is established in industrial practice – due to the following publications: Kurbel 2016, Stadtler 2012, Stadtler and Fleischmann 2012, Vollmann 2010 for example. As described in Hax and Meal (1975) as well as in Drexl et al. (1994), the PPC consists of a hierarchy of the planning levels; see also Claus et al. 2015: aggregate production planning, master production scheduling, lot sizing and scheduling (see Figure 1). Rough production plans at the highest level over a long planning period are increasingly refined over the individual planning levels, with increasingly shorter planning periods being considered. Finally, for a short-term period in the immediate future, it is determined for each individual resource (mainly machines) which product or workpiece is processed with this resource from when to when. Thus, each level is linked to the previous one in the hierarchy. If the planning levels are specified and solved by linear optimization problems, the fulfillment of the result of one planning level is guaranteed by restrictions on the next planning level.



**Figure 1:** Concept of hierarchical production planning

In detail, see Claus et al. 2015, aggregate production planning performs capacity planning over a medium-term period for a production site and product types in such a way that the costs for additional capacity, inventories and eventually more (like transport costs) are minimal. Master production scheduling synchronizes the available and required capacities for end products and their important components across all production segments of the production site in such a way that the same costs as with aggregate production planning are minimal, and so that the production program determined by the aggregate production planning is fulfilled. The next planning level(s) depend on the production types like the ones shown in Figure 1. Primarily for job shops

the next two levels consists of lot sizing and scheduling. Lot sizing, multi-level lot sizing to be precise, determines production orders for end products and their components as well as individual parts (according to the bill of materials) for production segments in such a way that setup and inventory costs are minimal and so that the production program determined by the master production scheduling is fulfilled. Scheduling dispatches the operations from the work plans of the production orders to individual resources such as machines or employees so that the start and end dates of the production orders, determined by lot sizing, are met. In the case of flow shops a simultaneous lot sizing and scheduling is performed – on only one planning level.

In classical approaches for the PPC, economic objectives such as meeting due dates, i.e. avoiding delays, are regarded. In this case, the ecological dimension is usually not taken into account, resulting in too much waste and emissions (e.g. more frequent set-ups can lead to higher energy consumption). Therefore the hierarchical production planning has to be extended by suitable indicators.

Within the ecological oriented research, indicators for the condition and the influence on the environment have been developed. However, since the ecological footprint of manufacturing has many different characteristics, it cannot be represented in one key figure alone. Many standardizing guidelines have taken on the task of defining various indicators. As suggested in Zarte et al. (2019), this article is oriented on the GRI (Global Reporting Initiative) Standard. The GRI is an international and independent organization, which provide definitions, requirements and indicators for the three pillars of sustainability. The corresponding GRI standards represent global best practice for public reporting on sustainable impacts. Sustainability reporting based on these standards provides information on an organization's contributions to sustainable development, the associated impacts and how these are managed. In the industry, reporting according to the GRI standard is well established, e.g. by German automotive concerns, and is therefore of particular relevance. Regarding the ecological dimension, there are the following eight standards: materials, energy, water and effluents, biodiversity, emissions, waste, environmental compliance and supplier environmental assessment (see GRI 2020). However, the standards environmental compliance and supplier environmental assessment cannot be influenced by the PPC. The remaining standards include indicators such as the organization's energy consumption, greenhouse gas emissions and waste volumes.

### 3. Methodology of literature review

We follow the review methodology published in Vom Brocke et al. (2009). Their approach consist of five steps: (I) Definition of review scope, (II) conceptualization of

topic, (III) literature search, (IV) literature analysis and synthesis, (V) research agenda.

This article is limited to the steps (I) to (III) and steps (IV) and (V) can be proceed with the analysis tool described in section 4.

For the first step, i.e. review scope, we use the results of section 2.

For the second step, we follow the advice in Vom Brocke et al. (2009) to directly start the process with database search by keywords. As database we use ScienceDirect; we discuss their relevance later.

In order to obtain a wide range of potentially relevant hits in the first search process, our keywords are very generic. Based on the considerations in section 2, we derive the keywords for sustainability from the GRI standard. They are:

- biodiversity,
- CO<sub>2</sub>,
- ecological,
- emission,
- energy,
- environment,
- material,
- sustainable,
- waste and
- water.

In case of planning we use the main planning problems – and consider in addition their synonyms:

- aggregate production planning,
- supply network planning,
- master production scheduling,
- master production planning,
- lot sizing and
- scheduling.

In order to perform a literature search of ecological sustainability in hierarchical production planning, we perform a concrete keyword search as follows. At least one sustainable related keyword is linked with at least one planning level related keyword via the Boolean operator “AND”.

After the keyword search, the resulting articles are further evaluated if they are relevant in terms of the scope determined in section 2. For this purpose all abstracts and, if this is not sufficient, the full texts of the articles are analysed.

In addition, we add all articles (in our database), which are referenced by the already identified articles and which are relevant (for our search); we call this backward search.

ScienceDirect includes more than 16 million research articles of more than 2 500 journals. There are several guides which assess the quality of journals – mainly in terms of how often their articles are cited; e.g. British Academic Journal Quality Guide or the German VHB-Jourqual3 ranking.

We take a closer look at the VHB-Jourqual3 ranking, limiting ourselves to the ranks from A+ to B and looking at all 923 items in our database. We did not find any relevant articles in journals such as Journal of Scheduling. This is probably because very specific procedures and problems are published in this journal. Moreover, this might be the reason why we found only 6 journals from the (restricted) VHB-Jourqual3 ranking, each of which published at least 10 articles out of our 923 found articles and which are not searched by ScienceDirect. To give the reader some insights, these 6 journals are listed in the following – their VHB-Jourqual3 rank is given in parentheses and they are ordered by the number of the articles found: International Journal of Production Research (B), Computers and Industrial Engineering (B), Computers and Operations Research (B), Operations Research (A+) and Journal of the Operational Research Society (B). To the journals OR Spectrum (A), which contains 8 out of our 923 articles, and Production and Operations Management (A), which contains 3 out of our 923 articles, the consideration just mentioned probably is true, but these journals might become relevant in the future. Examples of journals searched by ScienceDirect are, again the VHB-Jourqual3 rank is given in parentheses: European Journal of Operational Research (A), International Journal of Production Economics (B) and Journal of Cleaner Production (B). Concrete 4 journals with VHB-Jourqual3 rank of A and 16 journals with VHB-Jourqual3 rank of B are searched by ScienceDirect.

#### 4. Literature database and analysis tool

The initial keyword search resulted in 4 039 hits. After the two-step evaluating process, we received 253 relevant articles. The subsequent backward search resulted in further 670 relevant articles. Thus, we stored 923 articles in our database.

All articles were analysed in terms of the problems addressed, modelling techniques, algorithms and empirical studies. This lead to the following categories at the highest cataloguing level: planning level, objective function and restrictions, modelling and solution approach, shop floor characteristics as well as numerical examples and results.

We use this database to obtain the relevant literature for specific research questions and also for a literature review in a publication. To support these two tasks, we have implemented an analysis tool. Our analysis tool can be used to narrow down the subject areas to focus on special topics regarding ecological sustainability in PPC. Free of charge access is possible via the following

website of our research group: [www.oth-regensburg.de/sustainable-production-planning](http://www.oth-regensburg.de/sustainable-production-planning).

In principle, the analysis tool is a search engine through our database. The search is controlled by:

- enhanced cataloguing as a refinement of the mentioned highest cataloguing level and
- natural language queries, which are interpreted by the tool in a logical way.

Technically, a SQL query is generated on the website via the programming language PHP. The SQL queries contain the search topics of the user. Then, a connection with the database is established, the query is processed and the results are shown on the website. Note that the usage of the analysis tool is described on the website.

The user of the analysis tool could ask for articles on energy consumption. Of course the database contains many articles on this criterion, but only a small selection is listed. With increasingly specific questions, the selection is narrowed down further and further. Thus, if the user is interested in

- single machine,
- energy consumption costs,
- tardiness and earliness is minimized simultaneously,
- Mixed Integer Non Linear Problem,
- heuristic solution procedure and
- JIT penalty

just a few articles are in our database, especially Lee et al. (2017).

## 5. Summary and outlook

This article indicates that research on hierarchical production planning is increasingly trying to improve sustainability. Through a very broad search, we identified 923 articles which address problems, questions, models, solutions etc. on ecology and simultaneously on hierarchical production planning. With a digital search only, not all journals can (inevitably) have been considered. By analyzing the articles referenced by those that are found via ScienceDirect (i.e. our backward search) it might be that we have reduced the limitations of ScienceDirect significantly. Thus, the 923 articles might include at least one article from each researcher who has published in this area. But, the coverage of the articles is probably not complete. All references to these 923 articles are available free of charge via our analysis tool on a website of the research group. Our tests show that our analysis tool selects a small number of very relevant articles on specific ecological research issues in production planning. We will use this capability of our tool to compare our research with the state of the art.

## References

Absi, N.; Dazère-Pères, S.; Kedad-Sidhoum, S.; Penz, B.; Rapine, C. (2013): Lot sizing with carbon

emission constraints, *European Journal of Operational Research*, Vol. 227(1), pp. 55-61.

- Akbar, M.; Irohara, T. (2018): Scheduling for sustainable manufacturing: A review, *Journal of cleaner production*, Vol. 205, pp. 866–883.
- Battini, D.; Persona, A.; Sgarbossa, F. (2014): A sustainable EOQ model: theoretical formulation and applications, *International Journal of Production Economics*, Vol. 149, pp. 145-153.
- Bouchery, Y.; Ghaffari, A.; Jemali, Z.; Dallery, Y. (2012): Including sustainability criteria into inventory models, *European Journal of Operational Research*, Vol. 222(2), pp. 229-240.
- Brundtland, G.H.; Khalid, M.; Agnelli, S.; Al-Athel, S.; Chidzero, B.J.N.Y. (1987): *Our common future*, World Commission on Environment and Development. New York.
- Claus, T.; Herrmann, F.; Manitz, M. (2015): *Produktionsplanung und -steuerung: Forschungsansätze, Methoden und deren Anwendungen*, Springer, Berlin, Heidelberg.
- Destatis (Statistisches Bundesamt) (2020): *Umwelt: Abfallbilanz (Abfallaufkommen/-verbleib, Abfallintensität, Abfallaufkommen nach Wirtschaftszweigen) 2018*.
- Drexel, A.; Fleischmann, B.; Günther, H.O.; Stadler, H.; Tempelmeier, H. (1994): *Konzeptionelle Grundlagen kapazitätsorientierter PPS-Systeme [Conceptual Foundations of Capacity-Oriented Systems for Production Planning and Control]*, in: *Zeitschrift für Betriebswirtschaftliche Forschung* 46: S. 1022–1045.
- Eurostat (European Statistical Office) (2017): *Water use by economic sector - public water supply, 2015 (million m<sup>3</sup>)*.
- GRI (2020): *Global Reporting Initiative: Standards*. <https://www.globalreporting.org/standards/>. Last access on 3rd December 2020.
- Hax, A.C.; Meal, H.C. (1975): *Hierarchical Integration of Production Planning and Scheduling*, in: *Logistics*, Vol. 1, *Studies in Management Sciences*, edited by Murray A. Geisler, Amsterdam: North Holland.
- IEA (2020), *Tracking Industry 2020*, IEA, Paris. <https://www.iea.org/reports/tracking-industry-2020>. Last access on 3rd December 2020.
- Kurbel, K. (2016): *Enterprise Resource Planning and Supply Chain Management - Functions, Business Processes and Software for Manufacturing Companies*. Berlin: Springer.
- Lee, S.; Do Chung, B.; Jeon, H.W.; Chang, J. (2017): A dynamic control approach for energy-efficient production scheduling on a single machine under time-varying electricity pricing, *Journal of Cleaner Production*, Vol. 165, pp. 552-563.
- Stadler, H. (2012): *Master Planning - Supply Network Planning*, in: Stadler, H.; Fleischmann, B.; Grunow, M.; Meyr, H.; Sürie, C. (eds.): *Advanced Planning in Supply Chains*, Springer, Berlin, Heidelberg, pp. 109-148.



- Stadtler, H.; Fleischmann, B. (2012): Hierarchical Planning and the Supply Chain Planning. Matrix, in: Stadtler, H.; Fleischmann, B.; Grunow, M.; Meyr, H.; Sürie, C. (eds.): Advanced Planning in Supply Chains, Springer, Berlin, Heidelberg, pp. 21-34.
- U.S. Environmental Protection Agency (2020): Sustainable Manufacturing. <https://www.epa.gov/sustainability/sustainable-manufacturing>. Last access on 3rd December 2020.
- Vollmann, T.E. (2010): Manufacturing Planning and Control Systems for Supply Chain Management (6. edition). New York, London: McGraw-Hill.
- Vom Brocke, J.; Simons, A.; Niehaves, B.; Niehaves, B.; Reimer, K.; Plattfaut, R.; Cleven, A. (2009): Reconstructing the giant: On the importance of rigour in documenting the literature search process, in: Newell, S.; Whitley, E.A.; Pouloudi, N.; Wareham, J.; Mathiassen, L. (Eds.): Proceedings of the 17th European conference on information systems (ECIS2009).
- Zarte, M.; Pechmann, A.; Nunes, I.L. (2019): Indicator framework for sustainable production planning and controlling, International Journal of Sustainable Engineering, Vol. 12, pp. 149-158.

#### **AUTHOR BIOGRAPHIES**

**Hajo Terbrack M.Sc., MBA** is a PhD student at the Chair of Production Economy and Information Technology at the International Institute (IHI) Zittau, a central academic unit of Technische Universität (TU) Dresden.

**Ingo Frank M.A.** is a PhD student at the Ostbayerische Technische Hochschule Regensburg.

**Professor Dr. Frank Herrmann** was born in Münster, Germany and went to the RWTH Aachen, where he studied computer science and obtained his degree in 1989. During his time with the Fraunhofer Institute IITB in Karlsruhe he obtained his PhD in 1996 about scheduling problems. From 1996 until 2003 he worked for SAP AG on various jobs, at the last as director. In 2003 he became Professor for Production Logistics at the University of Applied Sciences in Regensburg. His research topics are planning algorithms, optimisation and simulation for operative production planning and control. He is head of the Innovation and Competence Centre for Production Logistics and Factory Planning (IPF).

**Professor Dr. Thorsten Claus** holds the Chair of Production Economy and Information Technology at the International Institute (IHI) Zittau, a central academic unit of Technische Universität (TU) Dresden, and he is the director of the IHI Zittau.

**Marco Trost M.A.** is a PhD student and research associate at the Chair of Production Economy and Information Technology at the International Institute (IHI) Zittau, a central academic unit of Technische Universität (TU) Dresden.

**Matthias Götz M.Sc.** is a PhD student at the Chair of Production Economy and Information Technology at the International Institute (IHI) Zittau, a central academic unit of Technische Universität (TU) Dresden.

# Implementation of the optimizer of SOA system deployment architecture

A. P. Woźniak

Military University of Technology

Institute of Computer and Information Systems

Ul. Gen. Sylwestra Kaliskiego 2, 00-908 Warsaw, Poland

E-mail: [adrian.wozniak@wat.edu.pl](mailto:adrian.wozniak@wat.edu.pl)

## KEYWORDS

Service-Oriented Architecture, SOA, business process, optimization, simulation

## ABSTRACT

Optimization of business processes in SOA systems has been done using three separate types of methods: Resource Allocation, Service Scheduling and Service Composition. All three may influence each other, so the new method has been proposed to find an optimal combination of those three. It is based on a genetic algorithm that uses a simulator of the SOA system to evaluate solutions. The article describes a model for the optimization criteria for such solutions. Subsequently, some basic concepts used to implement the simulator and optimizer have been presented. Finally, the performance results of the optimizer have been described, including the conclusions on how they might be improved.

## INTRODUCTION

The optimization of the system performance has always been important. It is due to many reasons, but first and foremost due to the limitation of resources or drive to increase system performance. It is no different in the case of the Service-Oriented Architecture (SOA) systems. Nonetheless, there are a few differences in how the systems might be optimized. The differences are mainly caused by fragmentation of the SOA systems. To add value to such systems, many components must cooperate. Each component is a software module that may be implemented in a technology different than other components and deployed independently. The most important part of the SOA components is that they deliver services. A service is a function of a component that is usually provided through a *www*. Users get value out of the SOA system by invoking the so-called composite services or business processes which are sequences of services realized by components.

The literature includes 3 types of methods for optimizing business processes in the SOA systems: Service Composition, Service Scheduling and Resource Allocation. Each of them is focused on a different stage of the SOA system implementation or execution. The first type is the Resource Allocation. It consists in determining which components should be deployed on which servers. Each component may be deployed simultaneously on many servers. Therefore, during Resource Allocation, it is also decided how many component instances should be running. Example of such a method uses Quality of Service (QoS) constrains and

resource usage cost as an input (Almeida et al. 2006; Huang et al. 2016; Mennes et al. 2016). Then it searches for optimal allocation using Fixed Point Iteration technique. The second way of optimizing business processes in SOA is to use the Service Composition method, which is the most popular in the literature (e.g. Ebrahim 2011; da Silva et al. 2015; Zhao et al. 2017; Wang et al. 2011; Xianwen et al. 2009). This type of optimization method is used when a given service is available on multiple servers. Usually, it is because the component is deployed on many servers. The Service Composition method is about deciding which server should execute a service instance. Usage of genetic algorithm is very common in solving this problem. Example of such approach is presented by Ebrahim (Ebrahim 2011). He suggests using a genetic algorithm where the chromosome has a number of genes equal to the number of services that must be called in the process. Each gene indicates an instance of the service that should be called in the process. The best chromosomes are those that provide the best QoS with minimal cost of service and minimal diversity of suppliers. The third type of methods is Service Scheduling (Dyachuk and Deters 2008). It is executed last and it is least popular in the literature. It may be used when multiple service invocations are organized in a queue of one component. Then it is possible to determine the order of their execution. For example, the Service Scheduling method presented in (Dyachuk and Deters 2007) finds services on a critical path of a business process and prioritizes them in the component queue.

All of these three methods are considered independently in the literature, even though they influence each other. Different service composition methods may give best results on different allocations and service scheduling algorithms. It means that we should strive to optimize all three aspects. Such optimization concept is proposed in (Woźniak and Nowicki 2019). It is based on a SOA system simulator that is used to evaluate solutions. The simulator takes, as an input, the SOA system model, which includes: services, components, execution environments, servers, business processes, etc. Each of the above-mentioned elements should be described with attributes, such as a random variable resources (CPU and RAM) used by each service invocation. In addition to the model, the simulator takes, as an input, the matrix of resource allocation and selected algorithms of the Service Composition and Service Scheduling methods. During simulation, output values that constitute criteria for selecting the best solutions (out of those that were simulated) may be obtained. To search through solutions,

a genetic algorithm was used in combination with a brute force approach. It is a unique feature of the SOA system allowing to define the optimization criteria from the business process point of view, which will be subsequently translated into the infrastructure. The reason behind this is that in SOA, business processes may be mapped to services. The following optimization model is an extension of one presented in (Woźniak and Nowicki 2019). Its main difference from the three SOA optimization types of methods is that it takes into account that all of those three influences each other so it finds optimal three: resource allocation, service composition algorithm and service scheduling algorithm.

## OPTIMIZATION MODEL

Several optimization criteria for business processes in the SOA system may be defined. They are focused on two aspects of the process: service quality for the user and costs for the company. The first two criteria are the following:

1. Average execution time of business processes weighted by the expected number of instances of business process.

$$\bar{k}_1(t, X) = E(k_1(t, X))$$

where:

$$k_1(t, X) = \sum_{b=1}^B \left( \frac{H_b(t)}{\sum_{i=1}^B H_i(t)} \cdot \frac{\sum_{i=1}^{LR_b} CR_{i,b}(t, X)}{LR_b(t)} \right)$$

B – number of business processes,

t – simulation time,

X – analysed solution which consists: boolean matrix of allocation of components to servers (genotype), selected Service Scheduling and Service Composition algorithms,  $H_x(t)$  – expected value for the number of x-type business process instances running,

$CR_{i,b}(t, X)$  – random variable denoting the time of implementation of the i-th instance of the b-th business process during t,

$LR_b(t)$  – random variable denoting the number of completed instances of the bth business process during time t.

2. Average variance of execution time of business processes

$$k_2(t, X) = \sum_{b=1}^B \frac{H_b(t)}{\sum_{i=1}^B H_i(t)} \cdot \frac{\left( \frac{\sum_{i=1}^{LR_b} CR_{i,b}(t, X)}{LR_b(t)} - CR_{i,b}(t, X) \right)^2}{LR_b(t)}$$

The costs criteria are included in the form of resources that are used by the system to calculate how they should be minimized. They are defined in the following manner:

1. The expected amount of processor resources used to provide services over a given time t.

$$\bar{k}_3(t, X) = E(k_3(t, X))$$

where:

$$k_3(t, X) = \frac{\sum_{s=1}^S m_s tu_s(t, X)}{\sum_{s=1}^S (m_s \cdot Y(s)) \cdot t}$$

S – number of servers,

$m_s$  – computational power of the s-th server,

$tu_s(t, X)$  – random variable denoting the time spent by s-th server on processing services

$Y(s)$  – function Y (s) takes the value 1 if any component is assigned to server s.

2. expected utilization rate of allocated memory resources for the provision of services during t

$$\bar{k}_4(t, X) = E(k_4(t, X))$$

where:

$$k_4(t, X) = \sum_{s=1}^S (p_s \cdot Y(s)) \cdot t$$

$p_s$  – amount of RAM on s-th server.

However, during the tests of optimizers, the  $\bar{k}_3$  criterion had two effects. The first one led to maximizing the use of a processor, which was beneficial. The second effect promoted the solutions that had long queues leading to longer execution times of business processes, which was unintended. The effect was partly nullified by the first criterion. However, to increase the convergence of the method, it was decided to replace it with a simpler one:

The expected degree of utilization of allocated processor resources for the provision of services in a given time t

$$\bar{k}_5(t, X) = E(k_5(t, X))$$

where:

$$k_5(t, X) = \sum_{s=1}^S (m_s \cdot Y(s)) \cdot t$$

Furthermore, during the experiments, it was noticeable that some solutions with high evaluation scores, according to the above criteria, had many unrealized business processes. The process may be abandoned if a server does not have enough resources to realize it or is damaged. To solve that, the another criterion was added:

$$\bar{k}_6(t, X) = E(k_6(t, X))$$

where

$$k_6(t, X) = \sum_{b=1}^B r_b(t, X)$$

$r_b(t, X)$  – random variable denoting the number of unrealized instances of a b-th type process.

Therefore, the target function of such optimization may be defined as:

$$\bar{k}(t, X) = (\bar{k}_1(t, X), \bar{k}_2(t, X), \bar{k}_4(t, X), \bar{k}_5(t, X), \bar{k}_6(t, X))$$

Not all component allocations are acceptable solutions. At least three restrictions on the solution should be defined:

Restriction 1. Processor. The server processor power should exceed its consumption as resulting from the operations of the components and execution environments.

$$\forall_{s \in (1, S)} \sum_{k=1}^K K_{k,s} M_k + \sum_{su=1}^{SU} S_{su,s} MU_{su} < m_s$$

where:

K – number of components

$K_{k,s}$  – binary value whether the k component has been assigned to the server s

$M_k$  – amount of processing power consumed by k-th component

SU – number of execution environments,

$S_{su,s}$  – binary value whether the su-th execution environment has been assigned to the server s,

$MU_k$  – amount of processing power consumed by su-th execution environment.

Restriction 2. RAM. The server RAM resources should exceed its consumption as resulting from the operations of the components and execution environments.

$$\forall_{s \in (1, S)} \sum_{k=1}^K K_{k,s} P_k + \sum_{su=1}^{SU} S_{su,s} PU_{su} < p_s$$

where:

$P_k$  – amount of RAM consumed by k-th component

$PU_{su}$  – amount of RAM consumed by su-th execution environment

Restriction 3. At least one instance of each component should be deployed.

$$\forall_{k \in (1, K)} \exists_{s \in (1, S)} K_{k,s} = 1$$

## SIMULATOR IMPLEMENTATION

To evaluate the solutions using the above criteria, the simulator of the SOA system was implemented in Java

language with the DISSim library. DISSim is a toolset for developers that allows to perform discrete event simulation. It uses BasicSimObject class for every element that is simulated and generates events in simulation time. The second class is BasicSimEvent that is put on simulation calendar. A simulation engine searches through the calendar and picks up the nearest event. Every event has a code attached thereto that is executed when the simulation time comes.

### Class Model of the Simulator

The BasicSimEvent and BasicSimObject classes are abstract. The latter is a parent object to the Organization and Server classes (Figure 1). The organization starts and contains instances of business processes that are specified by the Business Process Definition. The business process is defined as a graph, in which each step is a service. All services in the process are interconnected by arches described with the probability of choosing each path (which reflects the operation of the gates in BPMN). The services are described by the processor power, the RAM they need for their operations and the volume of data that should be sent through the network to provide the service. The services are associated with the components that execute them. The components contain a list of execution environments on which they can run. The components and execution environments are run on a server, which is a simulation object. They are described by the processor power and RAM memory necessary for their operations. The servers have specified amount of the processing power and RAM needed to run the components and execution environments as well as to provide the services. In addition, the servers are described with a matrix of network bandwidth between them. What is more, each server class object contains random variables that indicate the time of its damage and repair. Server damage events are created after starting the simulator in random time according to the random variable assigned to the server. When a server damage event occurs, the server's status is changed to inoperative and a server repair event is generated at a random time from the time of failure. The repair event generates the damage event, etc.

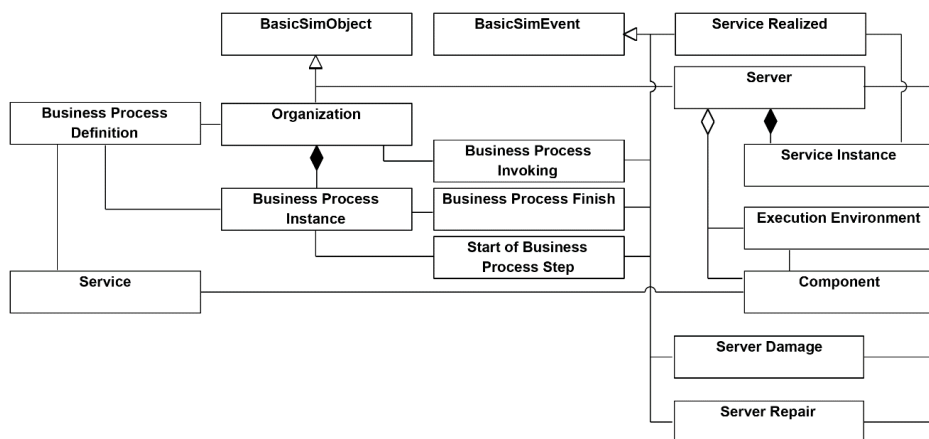


Figure 1 Class Model of the Simulator

## Simulation Process

### Creating the Business Process Instance

The events in the simulator are interdependent, and the logic of their occurrence is presented in Figure 2. After running the simulation, the "cyclic invocation of business process instance" events are generated. There is one event for every business process definition. It represents the creation of a new business process instance.

Once the simulation time reaches the event time, a new process instance and a new calendar event are created and will occur for the randomly generated simulation time. The time between the successive process start events is established according to the random variable specified in the business process definition.

### Business Process Realization

When the business process instance is created, its first step with the current simulation time is generated. Each event representing a step in the business process aims at invoking the services to accomplish such step. First of all, a server is appointed to execute the service. It is the operation of the load balancer, which consists in the selection of the correct instance of the component, i.e. implementation of the Service composition algorithm. Two Service Selection strategies have been implemented in the simulator:

- select least loaded server,
- select server with the shortest expected response time.

It is possible to add further Service Composition algorithms to the simulator. If a server capable of performing the step in the process is not found, then the process is terminated and the information about the

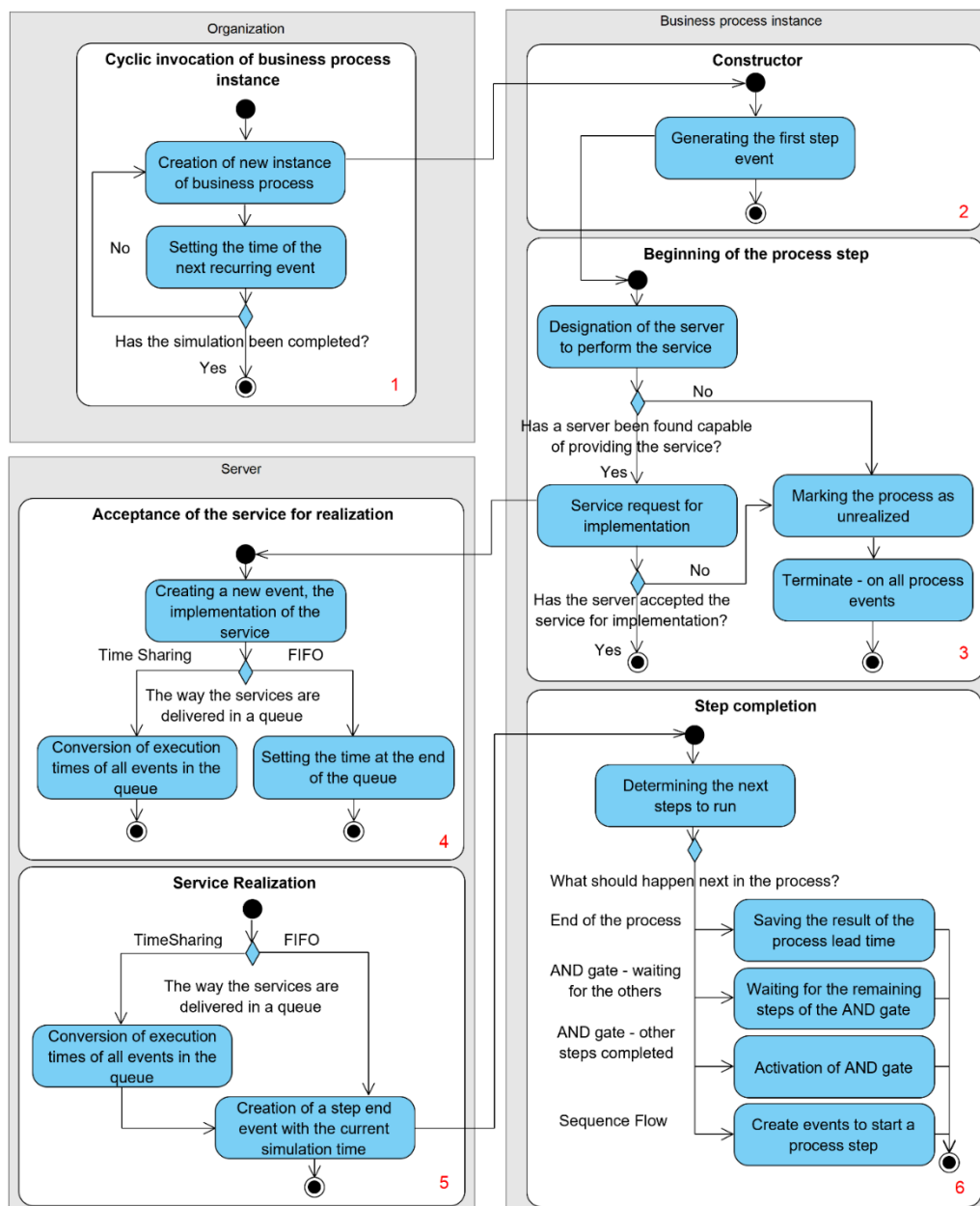


Figure 2 Simulation Process of the SOA System

system's inability to implement the process is included in the simulation results.

#### *Service Execution*

If the load balancer has found the server capable of providing the service, i.e. the server that:

- has enough free RAM memory,
  - has a component capable of providing the service,
- a service order is created. The time necessary for the service to be completed is the sum of:
- data transfer through the network,
  - service execution time on the server.

The transfer time depends on the volume of data to be transferred, as defined in the service, and the network bandwidth. The execution time depends on:

- the processor power allocated by the server to execute the services,
- the number of services invoking the orders,
- the power needed to perform the services,
- the component operation model (FIFO or Time Sharing).

If the component operates in the time-sharing mode, then each appearance of a new service to be executed and each termination of the service require recalculation of the expected service realization times.

#### *Business Process Step Execution*

The completion of the service creates a step completion event in the business process. Its aim is to determine the next steps. It may cause the termination of the process or creation of start events for one or more steps. If one of the XOR, OR or AND gateways was used after the step in the business process, then such step is interrelated with many other subsequent steps. Each relationship is described by the probability of path selection for XOR and OR gates. In case of the XOR gate, exactly one next step in the process is selected within the probability of the relations that add up to 1. In case of the OR gate, the probability of each path is calculated independently. Their sum may be greater than 1, hence, one or more of the following steps may be chosen. For the AND gate, all of the following steps are always run. The same is true for connecting gates (before the step) - in this case, it is essential to first complete one or more of the previous steps before starting the next step.

The above-described process is executed for every instance of the business process that can occur in large numbers. Simulation is performed for a fixed simulation time defined as a parameter before its beginning.

## **OPTIMIZER IMPLEMENTATION**

### **Population Initiation**

The optimizer is based on multiple simulations organized in a genetic algorithm and the brute force algorithm. The brute force algorithm is a loop that executes the genetic algorithm for each combination of the Service Composition and Service Scheduling methods. The genetic algorithm is used to find optimal resource allocation for the Service Composition and Service

Scheduling methods. The genetic algorithm starts with the generation of the population of genotypes. A genotype is a Boolean matrix that shows allocation of components to servers, where 1 indicates that the component has been allocated to the server. To generate an initial population, two layers of randomness are applied. The first one is the generation of numbers between 0 and 1 for each genotype. The number represents the probability of allocation. The second layer is randomization of 0 or 1, which shows whether the component is allocated to the server. This randomization is done with the probability of allocation from the first layer. In this way, not only different allocations are analyzed, but also solutions with different density of allocation.

After the generation of the initial population, it is necessary to adjust the solutions so that all restrictions are met. To that end, the genotypes that do not meet the first two restrictions are selected. Subsequently, the loop is executed for each server that is overloaded. Within each iteration of the loop, the randomly selected components and their execution environments are removed from the server allocation. The process continues until the server is not overloaded anymore. To guarantee that the third restriction is met, all genotypes that have at least one component not allocated to any server are selected. Thereupon, in case of each such component, a randomly selected server having enough resources to handle it is allocated.

### **Solution Evaluation and Selection**

To evaluate the solutions, each genotype is simulated. Each genotype is simulated multiple times to minimize the influence of randomness. During the simulation, the values of every decision criterium are subject to measurement. The criterion evaluation is averaged across multiple simulations of the same solution. After all solutions within a given population have been simulated, the ideal solution with the best values in each criterion in the population is formulated. The ideal solution is hypothetical. It is used as reference to evaluate other solutions. The values of all solutions in all criteria are normalized, where 0 constitutes the criterion value of the ideal solution and 1 - the worst value ever found. Subsequently, the distance from a given solution to the ideal one is calculated using the Euclidean metric. Best solutions are those that are closest to the ideal solution. The next step is the selection process. To do that, genotypes are sorted from best to worst. Survival chances are allocated to all solutions linearly, where the best solution has the probability of survival to the next population equal to 1. The worst solution has the probability of 0. The solutions are eliminated from the population according to the probability assigned thereto.

### **Crossover and Mutation**

The final steps of the genetic algorithm are crossover and mutation. During the crossover, new solutions are generated. Each new genotype has two parents randomly

selected. The probability of being selected is higher in case of better solutions. The weight of being picked as a parent is the same as the probability of survival in the previous step. Each pair of parents has two children. The genes of the children are randomly picked from one of the parents. If the first child inherited a gene from one parent, then the second child inherits it from the other parent. Next, new genotypes are undergoing mutation. There is a small probability that each gene will be mutated. Mutation is changing the value of a gene from 0 to 1 or from 1 to 0. This is to broaden the spectrum of the solutions searched. Finally, to ensure that all restrictions on solutions are met, the same algorithm as in case of the population initiation is performed.

## PERFORMANCE

The presented optimizer was used to find solutions to multiple problems. To test convergence, the same problem was optimized with multiple times using different seeds. Convergence depends on the following parameters:

- number of genetic algorithm iterations,
- size of population,
- variance of the simulation output data (which may be minimized by increasing the simulation time and number of repetitions),
- size of the problem (number of: servers, components, business processes, etc.).

The values of such parameters may be increased to achieve better convergence, but it would also make the optimization time longer. In the end it all comes down to the processing power and time. The more we have of those, the better convergence may be achieved.

The simulation time is not only dependent on how long it has to be processed, but also on how many events must be executed in the environment. The number of events depends on the number of business processes and their two attributes:

- expected value of a random variable of time between business process invocations,
- expected number of steps in a process to complete it.

Furthermore, there is one more value that has great impact on the simulation time. It is the ratio of load generated by the processes to available resources. The more load generated by the processes in comparison with the available resources, the longer service queues on the components. When the service instance is executed, the estimated execution times of all other service instances are updated. In case of a long queue, a lot of services have to be updated. According to the data gathered by a Java profiler, the service instance updating the process may consume up to 90% of the computation power provided to the optimizer. In case of a very short queue (shorter than 1 on average), it does not consume so many resources and the simulation process may be performed up to 10 times faster. Additionally, during experiments, it turned out that with short queues, the Service Scheduling algorithm of the optimal solution had a very low convergence. It is almost as if it was selected randomly. The reason for this is that when the queue is

short, then the Service Scheduling algorithm has nothing to optimize.

Table 1 shows the execution times on different parameters for the problem that comprises: 100 business processes and an average of 17.5 steps needed for their execution. Each business process definition had an average time between launches in a range of 1 to 500. This average time was a parameter for calling subsequent instances of business processes with exponential distribution. However this average time between business process instances was subject to changes in different simulation variants. The services were assigned an average execution time on a standard processor from 5 to 120 (note that one server has multiple processors so real execution time can be much shorter). This value was an input parameter for the execution time of individual service instances that were randomized according to the normal distribution (both the average time and standard deviation). The simulation length was 2000 time units. Every solution was simulated 10 times to evaluate the values of its decision criteria. There was  $3 \cdot 10^6$  solutions searched. Each solution had resource allocation problem size of 50 components allocated to 30 servers. The ratio between the resources required by the processes and the resources available on servers were from 2 to 1 (note that not all servers are used to minimize the resources consumed in the optimal solution). Optimization has been done on virtual machine with 3 Intel Xeon E5-2640 2.60 GHz virtual cores.

Table 1. Optimization times of problems with different amount of business process instances

Average time between invocations	Average number of events in simulation	Total optimization time (days)	Average time of evaluation of 1 solution (seconds)
10	350000	15,202	0,438
20	175000	8,445	0,243
40	87500	6,110	0,176

The differences in values of decision criteria between the variants varied between 1% and 15%, except for the variance of the business process execution time, which differed between 18% and 47%. Best convergence (1% difference between best solutions of different seeds) had been achieved when optimization was performing for 60 days for a problem of same size as presented above.

Another experiment was also carried out on the same problem, but with different population sizes and the number of iterations as shown in table 2. Each experiment was repeated five times, and the results of the time of each repetition are shown in Figure 3. The results show that the number of iterations in the proposed solution has a much greater impact on simulation time than the size of the population. Experiment 4 was carried out much longer than experiment 3 despite the need to review the same number of solutions. The reason for this may be greater optimization convergence with the

parameters of experiment 3 and thus a greater proportion of simulations closer to optimal, which require less computing power. Concept to combine all three types of optimization methods of business processes in SOA is new so it is impossible to compare those results to others.

Table 2. Parameters of experiments

	Exp. 1	Exp. 2	Exp. 3	Exp. 4
Iterations	5 000	5 000	5 000	10 000
Population size	100	150	200	100

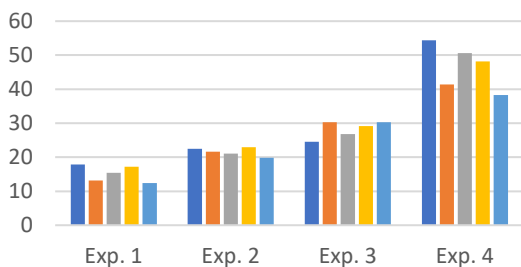


Figure 3 Execution time (days) of each experiment repetition.

## SUMMARY

The results of the proposed method may be interpreted dually. On one hand, it may give optimal solutions with high convergence, but on the other, it greatly depends on the resources and the problem size. To optimize it, greater utilization of multi-threading should be implemented in the simulator. Each simulation should be run as a separate thread. If further increase of its efficiency is needed, additional changes could be made. For example, each solution could be simulated on a different virtual machine. It is up to the user to decide if such efficiency of the optimizer is satisfactory. Still, the core of the optimizer would be same as described in the paper. What is more, with proper amount of time, it may have high convergence. As this method is new, it is impossible to compare it to other optimizers.

## REFERENCES

- Almeida J.; V. Almeida; D. Ardagna; C. Francalanci; and M. Trubian. 2006. "Resource Management in the Autonomic Service-Oriented Architecture". *IEEE International Conference on Autonomic Computing*. Dublin, Ireland.
- BPMN Specification documents. Accessed 31.01.2020. <https://www.omg.org/spec/BPMN/2.0/About-BPMN/>.
- da Silva A. S.; H. Ma; and M. Zhang. 2015. "A GP approach to QoS-aware web service composition including conditional constraints". *IEEE Congress on Evolutionary Computation (CEC)*. Sendai, Japan.
- Dyachuk D.; and R. Deters. 2008. "Ensuring Service Level Agreements for Service Workflows". *IEEE International Conference on Services Computing*. Honolulu, USA.
- Dyachuk D.; and R. Deters. 2007. "Service Level Agreement Aware Workflow Scheduling". *IEEE International Conference on Services Computing*. Salt Lake City, USA.

- Ebrahim G. A. 2011. "Intelligent Composition of Dynamic-Cost Services in Service-Oriented Architectures". *Fifth UKSim European Symposium*. Madrid, Spain.
- Huang K. C.; Y. C. Lu; M. H. Tsai; Y. J. Wu; and H. Y. Chang. 2016. "Performance-Efficient Service Deployment and Scheduling Methods for Composite Cloud Services". *IEEE/ACM 9th International Conference on Utility and Cloud Computing (UCC)*. Shanghai, China.
- Mennes R.; B. Spinnewyn; S. Latre; and J. F. Botero. 2016. "GRECO: A Distributed Genetic Algorithm for Reliable Application Placement in Hybrid Clouds". *5th IEEE International Conference on Cloud Networking (Cloudnet)*. Pisa, Italy.
- Schmid M. 2011. "An approach for autonomic performance management in SOA workflows". *12th IFIP/IEEE International Symposium on Integrated Network Management and Workshops*. Dublin, Ireland.
- Wang Z. J.; Z. Z. Liu; X. F. Zhou; and Y. S. Lou. 2011. "An approach for composite web service selection based on DGQoS". *The International Journal of Advanced Manufacturing Technology*, 56 (9). London, Great Britain 1167-1179.
- Woźniak A.; and T. Nowicki. 2019. "The Problem of Effective Deployment Architecture in SOA". *Computer Science and Mathematical Modelling* (9). Warsaw, Poland. 33-44.
- Xianwen F.; F. Xiaoqin; and C. Jiang. 2009. "An Efficient Approach to Web Service Selection". *Web Information Systems and Mining: International Conference*. Shanghai, China. 271-280.
- Xie L.; J. Luo; J. Qiu; J. A. Pershing; Y. Li; and Y. Chen. 2008. "Availability "weak point" analysis over an SOA deployment framework". *IEEE Network Operations and Management Symposium*. Salvador, Bahia, Brazil.
- Zhang C.; R. N. Chang; C. S. Perng; E. So; C. Tang; and T. Tao. 2009. "An Optimal Capacity Planning Algorithm for Provisioning Cluster-Based Failure-Resilient Composite Services". *IEEE International Conference on Services Computing*. Bangalore, India.
- Zhao Y.; W. Tan; and T. Jin. 2017. "QoS-aware Web Service Composition Considering the Constraints between Services". *12th Chinese Conference on Computer Supported Cooperative Work and Social Computing*. Chongqing, China.



**ADRIAN P. WOŹNIAK** studied at the Military University of Technology in Warsaw, where he obtained a degree in Information Technology in 2012. Subsequently, he worked as a system and business analyst. In this role, he designed many SOA systems. After that, his position changed to Architect and then Main IT Architect in the largest retail company in Poland, which allowed him to gain extensive experience in the field of the SOA system optimization problem. At the Military University of Technology, he is a lecturer in software engineering, system design and system integration. The subjects he teaches are also his main areas of research. His e-mail address is: [adrian.wozniak@wat.edu.pl](mailto:adrian.wozniak@wat.edu.pl)



# Evaluierung der Wahrnehmung von Chatbots im Kundenservice zur Optimierung der Mensch-Maschine-Interaktion

Daniela Mema

Hochschule Pforzheim  
Tiefenbronner Straße 65  
75175 Pforzheim  
Daniela.Mema@outlook.de

Prof. Dr. Véronique  
Goehlich

Hochschule Pforzheim  
Tiefenbronner Straße 65  
75175 Pforzheim  
Veronique.Goehlich@hs-  
pforzheim.de

Prof. Dr. Frank Morelli

Hochschule Pforzheim  
Tiefenbronner Straße 65  
75175 Pforzheim  
Frank.Morelli@hs-pforzheim.de

## ABSTRACT

Textbasierte Chatbots finden vermehrt Anwendung im Kundenservice der Mobilitätsbranche. Der vorliegende Beitrag gibt einen Überblick über die Anwendung von Chatbots im Kundenservice. Das Ziel der Forschung ist es, zu bestimmen, wie textbasierte Chatbots im Kundenservice von der deutschen und französischen Bevölkerung in der Mobilitätsbranche wahrgenommen werden. Hierbei werden in der Untersuchung die zwei größten Bahngesellschaften Deutschlands und Frankreichs als Beispiel herangezogen. Die Ergebnisse der quantitativen und qualitativen Forschung ergeben einige Zusammenhänge wie auch Unterschiede in der Wahrnehmung zwischen Deutschland und Frankreich. Die Kundenzufriedenheit sowie die Perzeption des Unternehmensimages beeinflusst die Wahrnehmung von textbasierten Chatbots im Kundenservice. Des Weiteren hängt die Empfindung gegenüber Chatbots mit der digitalen Affinität des Kunden zusammen. Zum aktuellen Zeitpunkt eignen sich Chatbots lediglich für die Beratung simpler Anliegen in den Augen deutscher und französischer Kunden. Der Funktionsumfang sowie das Design stellen weitere Einflussfaktoren für eine positive Wahrnehmung des Unternehmens-Chatbots dar.

## SCHLÜSSELWÖRTER

Chatbots, Kundenservice

## PROBLEMSTELLUNG UND

### VORGEHENSWEISE

Die vorliegende Studie greift eine aktuelle Thematik auf, die sich in der Entwicklung befindet. Eine weitgehende Erforschung der Wahrnehmung von Chatbots im Kundenservice der Mobilitätsbranche liegt zu diesem Zeitpunkt noch nicht vor. Darüber hinaus existiert in der Fachliteratur bis zum heutigen Zeitpunkt keine Gegenüberstellung der Wahrnehmung von Chatbots in der französischen und in der deutschen Gesellschaft. Im Folgenden ist von Chatbots in einem Business-to-Consumer (B2C) Umfeld auszugehen. Im Rahmen dieser Arbeit werden ausschließlich textbasierte Chatbots behandelt. Technische Aspekte werden in dieser Ausarbeitung demgegenüber nicht beleuchtet.

### *Zielsetzung und Forschungsfrage*

Das primäre Ziel der vorliegenden Arbeit besteht darin, die Kundenwahrnehmung von Chatbots im Kundenservice zu eruieren. Dabei wird auf die kulturellen Differenzen der französischen und deutschen Kunden in der Mobilitätsbranche eingegangen. Um die Wahrnehmung messen zu können, stellt die Kundenzufriedenheit ein relevantes Kriterium dar. Die Forschungsfrage des Artikels befasst sich demnach

damit, wie Chatbots im Kundenservice von deutschen und französischen Kunden wahrgenommen werden und in welchen Bereichen Optimierungsbedarf besteht. Die Relevanz der Forschungsfrage wird aus dem kontinuierlichen Wandel im Kundenservice ersichtlich. Dieser ergibt sich aus der Entwicklung von neuen Technologien und den wachsenden Kundenansprüchen im digitalen Zeitalter.

### *Vorstellung der empirischen Studie*

Die unterschiedlichen Wahrnehmungen von Chatbots im Kundenservice zwischen Deutschland und Frankreich können von Mentalitätsunterschieden der Kulturen geprägt sein. Aus diesem Grund erforscht der empirische Teil der Arbeit in welchen Bereichen diese Unterschiede existieren und in welchen Gebieten keine Unterschiede vorzufinden sind. Zum weiteren Vorgehen werden sieben wissenschaftliche Hypothesen gebildet. Diese bestehen jeweils aus mindestens zwei deskriptiven Variablen und weisen eine präzise Formulierung vor. Zur späteren Durchführung der inferenzstatistischen Tests werden jeweils eine Nullhypothese ( $H_0$ ) und ihre Alternativhypothese ( $H_1$ ) gebildet. Aufgrund der bisher wenig erforschten Thematik erfolgt die Entscheidung zu ungerichteten Alternativhypothesen ( $H_1: \mu \neq \mu_0$ ) (Hartmann und Lois, 2015, S.27). Im Folgenden werden die formulierten Nullhypothesen und Alternativhypothesen aufgezeigt. Die Alternativhypothesen gehen stets von einem Zusammenhang beziehungsweise von einem

Unterschied zwischen den jeweiligen Variablen aus, während die Nullhypothesen keinen Unterschied und keinen Zusammenhang beinhalten.

- **H0<sub>1</sub>/H1<sub>1</sub>**: Es gibt (k)einen Unterschied in der Kenntnis von Chatbots zwischen Deutschland und Frankreich.
- **H0<sub>2</sub>/H1<sub>2</sub>**: Es gibt (k)einen Unterschied in der Wahrnehmung von Chatbots zwischen Deutschland und Frankreich.
- **H0<sub>3</sub>/H1<sub>3</sub>**: Es gibt (k)einen Unterschied in der Wahrnehmung des Unternehmens-Chatbots zwischen Deutschland und Frankreich.
- **H0<sub>4</sub>/H1<sub>4</sub>**: Es gibt (k)einen Unterschied in der Wahrnehmung des Unternehmensimages durch den Einsatz von Chatbots zwischen Deutschland und Frankreich.
- **H0<sub>5</sub>/H1<sub>5</sub>**: Es gibt (k)einen Zusammenhang zwischen der digitalen Affinität des Nutzers und der Wahrnehmung von Chatbots im Kundenservice.
- **H0<sub>6</sub>/H1<sub>6</sub>**: Es gibt (k)einen Zusammenhang zwischen der Kundenzufriedenheit der SNCF/DB und der zukünftigen Nutzungsbereitschaft des Unternehmens Chatbots.
- **H0<sub>7</sub>/H1<sub>7</sub>**: Es gibt (k)einen Unterschied in der Präferenz der Kommunikationskanäle zur Interaktion mit Unternehmen zwischen Deutschland und Frankreich.

#### *Methodisches Vorgehen*

Als Erhebungsmethode wird das Mixed-Methods-Design angewandt. Diese Methode beinhaltet eine Einbettung quantitativer und qualitativer Methoden in das Forschungsprojekt. Mixed-Methods wird in ein- oder mehrphasigen Entwürfen angewendet (Kuckartz, 2014, S.33). In der vorliegenden Arbeit applizieren die Autoren die Forschungsmethode Mixed-Methods in zwei Phasen: Zu Beginn erfolgt die quantitative Forschung, um festzustellen, wie Chatbots im Kundenservice von Kunden oder potenziellen Kunden der Deutsche Bahn AG oder der SNCF wahrgenommen werden. Hierbei werden zwei inhaltlich nahezu identische Fragebögen in den Sprachen deutsch und französisch erstellt. Mithilfe der deutschsprachigen Fragebögen werden die Teilnehmer zu ihrer Wahrnehmung der Kundenkommunikation mit der Deutsche Bahn AG befragt. Demgegenüber erfolgt mithilfe der französischsprachigen Fragebögen die Befragung der Teilnehmer zu ihrer Wahrnehmung hinsichtlich der Kundenkommunikation mit der SNCF. Nach Beendigung und Auswertung der quantitativen Forschung wird eine qualitative Forschung realisiert. Hierfür erfolgt die Durchführung eines leitfadengestützten Interviews mit ausgewählten Personen. Den Interviewpartnern sollen beide Unternehmen, die Deutsche Bahn sowie die SNCF, bekannt sein. Durch die Multiperspektivität des ausgewählten Designs, wird die Forschungsfrage auf eine ausführliche Weise geprüft. Die Mixed-Methods

Forschung erfordert eine Kompatibilitätsannahme zwischen den beiden etablierten Forschungsmethoden (Kuckartz, 2014, S.35). Unter Berücksichtigung der Forschungsfrage sind in der vorliegenden Studie beide Methoden miteinander vereinbar. In Bezug auf den relativen Stellenwert der beiden Methoden in dem Mixed Methods-Projekt stellt die quantitative Forschung in der vorliegenden Arbeit die Kernkomponente dar. Die qualitative Forschung bildet somit die ergänzende Komponente (Kelle, 2019, S. 165f.), was man in der einschlägigen Literatur auch als explanatives Design bezeichnet (Hussy, Schreier, und Echterhoff, 2013, S. 299ff.). Durch diesen Methodenmix wird eine Überprüfung der aus der quantitativen Forschung gewonnenen Ergebnisse nochmals mithilfe der qualitativen Forschung durchgeführt.

#### **TRENDS IM KUNDENSERVICE**

##### *Kundenwahrnehmung*

Zur Eruierung der Wahrnehmungsunterschiede von Chatbots zwischen Deutschland und Frankreich werden im ersten Schritt die zusammenhängenden Wirkungsprozesse der Kundenwahrnehmung im Allgemeinen dargestellt. Die Kundenwahrnehmung stellt eine geschlossene Wirkungskette dar (Töpfer und Mann, 2008, S.43). Das langfristige Ziel einer verbesserten Wahrnehmung durch den Kunden ist die Kundenbindung. Die Wirkungskette beginnt mit der Analyse der Kundenbedürfnisse, welche weitgehend unkonkreter und produktspezifischer Natur sind. Die Untersuchung von Kundenbedürfnissen ergibt die existierenden Mangelgefühle des Abnehmerkreises. Diese primäre Phase beschäftigt sich demzufolge mit der Feststellung des Mangels und der lösungsorientierten Vorgehensweise mit dem Ziel der Beseitigung dieses Mangels (Trommsdorff, 2004 zitiert nach Töpfer und Mann, 2008, S. 42). In der sekundären Phase werden die konkreten Bedürfnisse festgestellt und als Kundenanforderungen formuliert. Im nächsten Schritt, der Analyse der Kundenerwartungen, werden die identifizierten Bedürfnisse auf eine konkrete Dienstleistung oder ein Produkt bezogen. Im Regelfall sind die Kundenerwartungen zusätzlich von einem Vergleich der Problemlösungen geprägt, die sich zu dem Zeitpunkt auf dem Markt befinden. Die Kundenwahrnehmung ist stets als subjektiv zu erfassen und inkludiert eine Abweichung von objektiven Faktoren. Aufgrund dieser Subjektivität verfügt die Kundenwahrnehmung über eine Filterfunktion, die über die weiteren Wirkungsstufen entscheidet (Bruhn, 2000 zitiert nach Töpfer und Mann, 2008, S. 42f.). Das Zufriedenheitsurteil des Kunden basiert auf seiner Wahrnehmung der Dienstleistung oder des Produktes. Bei einem positiven Ergebnis bildet die Kundenzufriedenheit die Basis für die Kundenloyalität und die Kundenbindung, welche eines der langfristigen Ziele eines Unternehmens in Bezug auf den Unternehmenserfolg darstellt (Töpfer und Mann, 2008, S. 43.). Abschließend lässt sich feststellen, dass die Kundenwahrnehmung mit der Kundenzufriedenheit

sowie den nachfolgenden Etappen innerhalb der Wirkungskette zusammenhängt. Aus diesem Grund ist es sinnvoll, die Kundenzufriedenheit im Detail zu analysieren, um auf die Kundenwahrnehmung von Chatbots Rückschlüsse ziehen zu können. Dies erfolgt im Rahmen der empirischen Studie.

#### *Kundenzufriedenheit*

Die Rolle des Kundenservice nimmt eine wachsende Bedeutung ein. Im heutigen Zeitalter genügt es für Unternehmen nicht, ein gutes Produkt einzuführen, um erfolgreich zu sein. Stattdessen wird von den Kunden ein umfangreiches Servicekonzept der Unternehmen erwartet. Um den Kunden langfristig zu binden und einen Mehrwert zu schaffen, antworten Unternehmen auf diese Anforderungen mit innovativen Servicestrukturen (Gouthier, 2016, S. 5f.). Die Gründe für die Abwanderung von Kunden sind vielseitig. Es lässt sich jedoch ein dominanter Grund feststellen, den 68 Prozent der abgewanderten Kunden in einer Umfrage der Canadian Management Association angeben. Für diesen Anteil der abgewanderten Kunden stellt die Unzufriedenheit im (Kunden-)Service einen ausschlaggebenden Abwanderungsgrund dar. Lediglich neun Prozent der Abwanderungen sind auf ein besseres Konkurrenzprodukt zurückzuführen. Darüber hinaus wandern fünf Prozent der Kunden aufgrund von Geschäftsbeziehungen mit befreundeten Unternehmen ab. Daraus lässt sich schließen, dass der Kundenservice eine Determinante der Kundenzufriedenheit und der Kundenbindung darstellt. Um einen zufriedenstellenden Kundenservice anbieten zu können, arbeiten Unternehmen daran, ihren Kundenservice zu optimieren (Töpfer und Mann, 2008, S.21). Die Qualitätsverbesserung und die Kostensenkung stellen Kontroversen dar, mit denen sich Unternehmen auseinandersetzen. Aufgrund der Markttransparenz ist es für Unternehmen erstrebenswert, ein Gleichgewicht zu schaffen und sich von anderen Wettbewerbern abzuheben (Gouthier, 2016, S.15). Darüber hinaus herrscht eine Wechselbeziehung zwischen der Kundenzufriedenheit und der wahrgenommenen Wertschätzung des Kunden. Entwickelt der Kunde das Gefühl, serienmäßig behandelt zu werden, sinkt die wahrgenommene Wertschätzung und folglich die Kundenzufriedenheit. Unternehmen stehen vor der Herausforderung darüber zu entscheiden, bis zu welchem Grad sie automatisierte Prozesse im Kundenservice einsetzen. Dabei stellt sich die Frage, wie diese von den Kunden wahrgenommen werden. Demgegenüber steigern standardisierte Prozesse die Effizienz und senken die Kosten (Gouthier, 2016, S.16).

#### *Digital Customer Experience*

Einen weiteren Einflussfaktor der Wahrnehmung von Chatbots im Kundenservice bildet die Digital Customer Experience, die der Kunde bei der Interaktion mit einem Unternehmens-Chatbot wahrnimmt. Das Begriffspaar Customer Experience leitet sich aus dem Wort „Customer“ (Kunde) und „Experience“ (Erfahrung,

Erlebnis) ab. Die Überführung des letzteren Begriffs ist nicht eindeutig und kann als Kundenerfahrung oder als Kundenerlebnis gedeutet werden. Das Kundenerlebnis beschreibt einen kurzfristigen, emotionalen Vorgang. Demgegenüber stellt die Kundenerfahrung eine Verkettung einzelner Kundenerlebnisse dar (Böhme & Potyka, 1995 zitiert nach Frohne, 2020, S. 12f.). Die Customer Experience umschließt sämtliche persönliche Wahrnehmungen und Interaktionen eines Kunden mit den unterschiedlichen Touchpoints eines Unternehmens. Aus diesem Grund ist sie von dem Konzept der Kundenzufriedenheit, das sich auf die Erfüllung der Erwartungen des Kunden stützt, abzugrenzen. Abweichend von dem Kundenzufriedenheitskonzept werden im Rahmen der Customer Experience nicht nur die positiven Erlebnisse des Kunden berücksichtigt, sondern die Gesamtheit aller Erlebnisse und Interaktionen (Schmitt, 2009 zitiert nach Frohne, 2020, S. 11). Die steigenden Anforderungen des Kunden an ein Einkaufserlebnis sowie substituierbare Produkte und Dienstleistungen verstärken den Wettbewerb. Vor diesem Hintergrund verfolgen Unternehmen die Erzeugung einer positiven Customer Experience (Holland, 2019). Die Digital Customer Experience beinhaltet ausschließlich Begegnungen des Kunden mit den digitalen Touchpoints eines Unternehmens oder dessen Marke, Produkten und Dienstleistungen. Mögliche digitale Schnittstellen sind Smartphones, Computer und Tablets. Gemäß einer Studie des Harvard Business Review ist die Toleranzgrenze der Kunden für Wartezeiten in einem Online-Umfeld geringer als in einem physischen Geschäft. Eine Wartezeit von zehn Sekunden für das Laden einer Website führt bei 50 Prozent zum Verlassen dieser Website. Bereits 90 Prozent der Kunden haben einen oder mehrere Kriterien zum Misserfolg in ihrer Digital Customer Experience erlebt bei der Benutzung von Mobilgeräten auf der Suche nach Informationen zum Kundenservice (Borowski, 2015).

#### *Omnichannel im Kundenservice*

In Anbetracht der wachsenden Kundenansprüche an eine stetige Erreichbarkeit des Kundenservice, gewinnt das Omnichannel Management an Bedeutung. Mobile Kommunikationskanäle werden in der heutigen Zeit von Kunden als selbstverständlich wahrgenommen. Unternehmen erweitern im Zuge dessen ihr Dienstleistungsspektrum mit dem Ziel, ein positives Kundenerlebnis zu gewährleisten (Verhoef et al., 2015 zitiert nach Robra-Bissantz und Lattemann, 2019, S. 41). Es wird zwischen traditionellen und digitalen Kommunikationskanälen unterschieden. Zu den traditionellen Kanälen zählen physische Kommunikationsformen nicht-digitalen Ursprungs. Aufgrund der Ubiquität des E-Mail-Verkehrs, wird diese Kommunikationsform jedoch gleichermaßen zu den traditionellen Formen gezählt. Digitale Kanäle weisen als Gemeinsamkeit die Nutzung durch das Internet auf (Straker et al., 2015 zitiert nach Robra-Bissantz und Lattemann, 2019, S. 41).

## CHATBOTS

Der Begriff Chatbot stammt aus dem Englischen und besteht aus den Wörtern „Bot“ (abgeleitet von „Robot“) und „Chat“ (plaudern) (Stucki, 2018, S. 3). Er bezeichnet ein Dialogsystem, welches auf natürlicher Sprache basiert (Bendel, 2019). Ein Chatbot antwortet dem Nutzer unter Anwendung von Datenauswertungen, Regeln und Routinen in Form eines natürlichen, textbasierten Satzgebildes. (Stanoevska-Slabeva et al., 2018 zitiert nach Stanoevska-Slabeva, 2018, S. 27ff.). Die Anwendung von Chatbots wird von Unternehmen oftmals mit einem Avatar kombiniert. Daraus ergibt sich eine markenspezifische Visualisierung des Chatbots. Es ist möglich, eine Persönlichkeit des Chatbots durch den Avatar aufzubauen. Die Betreuung von Chatbots findet über diverse Kommunikationsplattformen statt. Sie werden in die bestehenden Touchpoints des Unternehmens integriert. Unternehmen können ihren Chatbot über externe Plattformanbieter einsetzen. Des Weiteren ist es möglich, einen Chatbot auf der eigenen Unternehmenswebseite sowie auf einer mobilen Applikation anzubieten. Die Kommunikation mit Chatbots erfolgt über unterschiedliche Ausgabegeräte. Dazu zählen Smartphones und andere gängige Endgeräte wie Computer oder Tablets (Stanoevska-Slabeva, 2018, S. 26f.). Die zunehmende Verbreitung von Chatbots wird von dem kontinuierlichen technologischen Fortschritt und dem Gesellschaftswandel begünstigt. Durch die Verbindung der Forschungsbereiche der Künstlichen Intelligenz (KI) mit ihrem Teilgebiet des maschinellen Lernens entwickelt sich ein System von Chatbots. Dieses System ist in der Lage, autonom auf eine Aussage zu reagieren. Der heutige Stand der Chatbots beschreibt eine schwache KI, die als Teilintelligenz in einem konkreten Einsatzbereich verstanden wird (Kehl und Coenen, 2016 zitiert nach Rühle, Hoesch, und Petersohn, 2019, S.49). In der Wirtschaft werden Chatbots vermehrt in der Kundenkommunikation eingesetzt. Sie reagieren auf das Anliegen des Nutzers (Bitkom, 2017 zitiert nach Rühle et al., 2019, S. 49-50).

### *Status quo des Chatbot Markts*

Die durchschnittliche jährliche Wachstumsrate für den globalen Chatbot Markt wird gemäß des Marktforschungsunternehmens Grand View Research auf 24,3 Prozent für das Jahr 2025 geschätzt. Dieses signifikante Wachstum ist auf eine erhöhte Anzahl der Nutzer von Messaging-Diensten zurückzuführen. Darüber hinaus führt die stetige Entwicklung im Bereich Natural Language Processing zu einer fließenden Mensch-Maschine Interaktion. Daraus wird eine Steigerung der Marktnachfrage abgeleitet. Weiterhin wird der Chatbot-Markt nach ausschlaggebenden Kriterien segmentiert. Relevant ist das Gebiet der Anwendung von Chatbots. Weitere Anwendungsgebiete für Chatbots neben dem Kundenservice bestehen in den Bereichen der sozialen Medien, der Bezahlung und

Bestellabwicklung oder im Marketing. Dabei können sie Web-basiert, Messenger-basiert oder alleinstehend eingesetzt werden. Chatbots werden heutzutage in verschiedenen Bereichen verwendet. Dazu gehört das Gesundheitswesen, der Einzelhandel, das Bankwesen, Versicherungen, E-Commerce und auch die Reise- und Tourismusbranche (Grand View Research, 2017).

### *Anwendungsbereiche von Chatbots in der Kundenkommunikation*

In der Kundenkommunikation variieren die Anwendungsbereiche von Chatbots abhängig davon, welchen Mehrwert das Unternehmen durch den Einsatz des Chatbots erzielt. Chatbots lassen sich für beratende Kundenprozesse einsetzen, die auf grafischen Benutzerschnittstellen schwierig abbildbar sind. Sie bieten dem Kunden durch den Zugriff auf Datenquellen eine erhöhte Flexibilität und Individualisierung im Vergleich zu standardisierten Informationsquellen. Vor diesem Hintergrund nehmen Chatbots eine effiziente Funktion zur Produktberatung ein. Zudem stellen Chatbots eine Unterstützung für Kundendialoge dar, die mit dem Ausfüllen eines Formulars verbunden sind. Weitere Einsatzgebiete von Chatbots im Kundenservice sind Buchungsprozesse von Reisen oder Tickets. In diesem Fall wird ein Nutzen für Reisen mit umfassenden Kundenangaben geschaffen. Der Kunde wird von dem Chatbot entlang des Prozesses geführt und gelangt nach der Bezahlung direkt zu seiner Buchungsbestätigung. In diesem Zusammenhang stehen auch Transaktionsprozesse, die vom Chatbot eingeleitet werden. Einkaufs- und Zahlungsprozesse werden durch die Äußerung des Kunden über das Conversational Interface ausgeführt. Weiterhin werden Chatbots für komplexe Suchprozesse eingesetzt. Für den Fall, dass der Kunde zu bestimmten Informationen auf der Unternehmenswebseite nicht fündig wird, steht der Chatbot in einem Dialog zur Verfügung (Stanoevska-Slabeva, 2018, S. 32-34).

## KULTURDIMENSIONEN NACH GLOBE

Die beiden zu untersuchenden Länder Frankreich und Deutschland liegen geographisch in unmittelbarer Nähe. Gemäß des Global Leadership and Organizational Behavior Effectiveness Projekts (GLOBE) sind sie jedoch in unterschiedliche Kulturcluster einzuordnen. Die GLOBE-Kulturdimensionen basieren auf den Studien der Kulturwissenschaftler Geert Hofstede und Fons Trompenaars. Nach GLOBE existieren neun Kulturdimensionen. Die Untersuchung der Kulturdimensionen erfolgt nach zwei Kriterien: Es wird zwischen Praktiken und Werten differenziert. Die Praktiken spiegeln die tatsächlichen Erlebnisse in einer Gesellschaft wider, während die Werte die Meinung der Mitglieder, wie die Gesellschaft sein sollte, darstellen (House et al., 2004 zitiert nach Brodbeck, 2016, S. 73f.). Im Folgenden werden lediglich die für die Untersuchung relevanten Dimensionen erläutert.

### *Zukunftsorientierung*

Die Förderung von vorausschauenden Verhaltensweisen wird als Zukunftsorientierung definiert. Die Einordnung des Zeitbegriffs erfolgt nach der Vergangenheits- und Gegenwartsorientierung (Kluckhohn et al., 1961; Lewin K., 1942 zitiert nach Brodbeck, 2016, S. 98). Gesellschaften mit hohen Ergebnissen in der Dimension Zukunftsorientierung schätzen langfristige Planung und Geldanlagen. Sie bevorzugen Ersparnisse gegenüber impulsiven Einkäufen und sind geprägt von einer intrinsischen Motivation. In Gesellschaften mit niedrig ausgeprägter Zukunftsorientierung werden unmittelbare Investitionen gegenüber Ersparnissen bevorzugt. Die Werte der deutschen Kultur (4,85) und der französischen Kultur (4,96) liegen nah beieinander. Ein Unterschied wird jedoch in den Ergebnissen der Praktiken deutlich. Deutschlands Gesellschaft praktiziert eine höhere Zukunftsorientierung (4,27) als die Gesellschaft Frankreichs (3,48) (Ashkanasy et al., 2004 zitiert nach Brodbeck, 2016, S. 98-103). Diese Differenzen im Kaufverhalten erweisen sich als relevant in Bezug auf die Einsatzmöglichkeiten von Chatbots im Kundenservice.

### *Bestimmtheit*

Die kulturelle Dimension Bestimmtheit beschreibt das Verhalten der Mitglieder einer Gesellschaft gegenüber ihren Mitmenschen. Eine Gesellschaft weist eine hohe Ausprägung der kulturellen Dimension Bestimmtheit auf, wenn ihre Mitglieder bei der Interaktion mit Mitmenschen ein hohes Ausmaß an Direktheit oder Aggression zum Ausdruck bringen (Praktiken) oder zum Ausdruck bringen sollten (Werte). Die Kommunikation eigener Gedanken und Sinneswahrnehmungen verläuft eindeutig. Demgegenüber weisen Gesellschaften mit einer niedrigen Ausprägung eine Präferenz für Befürwortet Selbstbeherrschung, Solidarität und Tradition. Das Ergebnis der deutschen Gesellschaft (4,55) weist in Bezug auf die kulturellen Praktiken einen höheren Grad an Bestimmtheit auf als die französische Gesellschaft (4,13). Dahingegen weist die Kultur Frankreichs (3,38) einen höheren Wert als die Gesellschaft Deutschlands (3,09) auf. Daraus ist zu schließen, dass sich die Kommunikationsweise des Chatbots an den Grad der Bestimmtheit der jeweiligen Kultur adaptieren sollte (Den Hartog, 2004 zitiert nach Brodbeck 2016, S.108-112).

### *Gruppenbasierter Kollektivismus*

Die Kulturdimension des gruppenbasierten Kollektivismus bezeichnet den Grad, an dem das Wohlergehen einer nahestehenden Gruppe für ein Individuum an erster Stelle steht. In Gesellschaften mit einem hohen Ergebnis sind die Gruppenzugehörigkeit und insbesondere die Familienzugehörigkeit von großer Bedeutung. Es wird stark differenziert zwischen Gruppenmitgliedern und Individuen, die der Gruppe nicht angehören. Eine indirekte Kommunikation ist zwischen den Gruppenmitgliedern vorzufinden. In Gesellschaften mit niedrig ausgeprägtem Kollektivismus

dominiert der Individualismus. Ein rationales Weltbild sowie eine geringere Unterscheidung zwischen Mitgliedern und Nichtmitgliedern sind charakteristisch. Auffallend ist, dass französische Mitglieder der Gesellschaft in den kulturellen Praktiken (4,37) sowie in den kulturellen Werten (5,42) höhere Ergebnisse im Vergleich zu deutschen Praktiken (4,02) und Werten (5,18) aufweisen (Gelfand et al., 2004 zitiert nach Brodbeck 2016, S. 112-117). Diese Dimension ist aufgrund der Differenzen in der Praxis zwischen Deutschland und Frankreich maßgebend für die Wahrnehmung von Chatbots im Kundenservice.

### *Humanorientierung*

Die Kulturdimension Humanorientierung wird als das Ausmaß definiert, in dem bestimmte Verhaltensweisen gegenüber Mitmenschen gefördert und belohnt werden. Zu den beobachteten Verhaltensmustern gehören Höflichkeit, Großzügigkeit, Fürsorge und Fairness. In Gesellschaften mit einer hohen Ausprägung der kulturellen Dimension Humanorientierung werden die Mitglieder durch das Gefühl der Zugehörigkeit zu einer Gruppe motiviert. Des Weiteren schätzt man dort eine freundliche, großzügige Verhaltensweise. Im Gegenzug wird bei einer geringen Ausprägung der Humanorientierung ein hedonistischer Lebensstil bevorzugt. Damit geht eine komfortable, und auf Vergnügen ausgelegte Lebensführung einher. Der Staat nimmt eine unterstützende Position im Bereich Wirtschaft und Sozialwesen ein. Die Ergebnisse der GLOBE-Studie zeigen, dass Deutschland (5,46) und Frankreich (5,67) erhöhte kulturelle Werte für die kulturelle Dimension Humanorientierung erzielen. Die kulturellen Praktiken sind in Deutschland (3,18) niedriger anzusiedeln als in Frankreich (3,40). Die Praktiken dieser Dimension erweisen sich als relevant zur Berücksichtigung kultureller Differenzen in der Wahrnehmung von Chatbots (Kabasakal et al., 2004 zitiert nach Brodbeck 2016, S. 126-130).

### *Unsicherheitsvermeidung*

Die Kulturdimension Unsicherheitsvermeidung beschreibt die Intensität, in der traditionelle Verfahren und soziale Kontrolle genutzt werden. Ziel ist es, die Unvorhersehbarkeit von einkehrenden Ereignissen zu reduzieren. Gesellschaften mit einer hohen Ausprägung der kulturellen Dimension Unsicherheitsvermeidung bevorzugen klar definierte Regeln und Formalitäten. Die Kommunikation wird in diesen Kulturen schriftlich dokumentiert. Dabei sind bindende Verträge in einer schriftlichen Ausführung wesentlich. Gesellschaften mit niedrig ausgeprägter Unsicherheitsvermeidung umgehen Verhaltensregeln und bevorzugen eine informelle Kommunikationsweise. Neuen Entwicklungen von Produkten stehen Mitglieder dieser Gesellschaften meist offen gegenüber. In Bezug auf die kulturellen Werte erreicht Deutschland (3,32) einen niedrigeren Wert im Vergleich zu Frankreich (4,26). Dahingegen weist die deutsche Gesellschaft (5,22) ein höheres Ergebnis in der Praxis im Vergleich zu Frankreich (4,43) auf. Diese

Kulturdimension ist für die weitere Forschung dieser Arbeit relevant, da Chatbots neue Technologien darstellen und diese unterschiedlich von den beiden Kulturen aufgefasst werden (Sully de Luque et al., 2004 zitiert nach Brodbeck 2016, S. 130-136).

#### *Auswirkungen auf den Kundenservice in der Mobilitätsbranche*

Der heutige Kundenservice erfordert eine Sensibilisierung in Bezug auf die Gewohnheiten, Praktiken und Präferenzen des Kunden innerhalb seiner Kultur. Durch die Einbeziehung der kulturellen Dimensionen nach GLOBE werden die kulturellen Unterschiede deutlich. Nachfolgend werden diese hinsichtlich der Anwendung von Chatbots im Kundenservice der Mobilitätsbranche aufgeführt. Es ist hervorzuheben, dass die Gesellschaft Frankreichs erhöhte Werte der Dimension des gruppenbasierten Kollektivismus aufweist. Die Mitglieder erachten die Gruppenzugehörigkeit als wichtig. (Gelfand et al., 2004 zitiert nach Brodbeck 2016, S. 112-117). Im Kundenservice bedeutet dies die Präferenz eines einheitlichen Bildes des Kundenservices in Bezug auf die Anwendung der französischen Sprache. Der eingesetzte Chatbot sollte in französischer Sprache entwickelt sein und einen französischen Namen tragen, sodass er von den Mitgliedern der Gesellschaft akzeptiert wird. Die Dimension Unsicherheitsvermeidung ist relevant für den Kundenservice vor dem Hintergrund des fortschreitenden technologischen Fortschritts. Die Wahrnehmung von Chatbots in der Mobilitätsbranche kann durch diese Dimension genauer betrachtet werden. Die deutsche Gesellschaft weist erhöhte Ergebnisse in den Praktiken der Dimension Unsicherheitsvermeidung auf (Sully de Luque et al., 2004 zitiert nach Brodbeck 2016, S. 130-136). Dies bedeutet, dass die Mitglieder eine höhere Hemmschwelle gegenüber neuen Technologien wie Chatbots im Kundenservice haben. Im Vergleich dazu zeigt Frankreichs Gesellschaft niedrigere Resultate in Bezug auf die Unsicherheitsvermeidung auf. In diesem Fall wird von einer erhöhten Akzeptanz der Mensch-Maschine- Interaktion im Kundenservice ausgegangen. Weiterhin wird eine informelle Kommunikationsweise bevorzugt. Frankreich weist eine niedrigere Ausprägung der Zukunftsorientierung in der Praktik auf im Vergleich zu Deutschland (Ashkanasy et al., 2004 zitiert nach Brodbeck 2016, S. 96-133). Im Zusammenhang mit Chatbots sind die Differenzen in dieser Dimension relevant für das Kaufverhalten der Mitglieder. Daraus ist zu schließen, dass Mitglieder der französischen Gesellschaft eher dazu neigen, impulsive Einkäufe gegenüber Ersparnissen zu tätigen. Diese Präferenz ist in der Funktionsweise des Chatbots zu berücksichtigen, da damit weitere Verkaufspotenziale im Bereich des Cross- und Upsellings verbunden sind. Ferner ist eine erhöhte Ausprägung der Humanorientierung in der französischen Kultur festzustellen. Daraus ist zu schließen, dass eine höfliche und zuvorkommende Verhaltensweise des Chatbots

einen hohen Stellenwert in der französischen Gesellschaft hat.

#### **FALLBEISPIELE MOBILITÄTSBRANCHE**

##### *Status quo des Kundenservice der Deutsche Bahn AG*

Der Kundenservice des Personenverkehrs der Deutsche Bahn AG ist über unterschiedliche Kommunikationskanäle erreichbar. Es stehen fünf telefonische Auskunft -und Beratungsdienste bereit, die sich in ihrer Funktion unterscheiden. Der Kunde wählt über einen Anruf zur Auskunft über Fahrpreise und Fahrpläne sowie über Servicedienstleistungen, Fahrkartenbuchungen, der Planung barrierefreier Reisen bis hin zu einer Fundservice Hotline. Jeder der Telefondienste ist für den Kunden kostenpflichtig. Darüber hinaus ist es möglich, die sozialen Netzwerke Facebook und Twitter für die Kontaktaufnahme mit dem Unternehmen zu verwenden (Deutsche Bahn AG, o.D.).

##### *Chatbot Kai*

Der Service Chatbot der Deutsche Bahn AG ist über die Unternehmenswebseite erreichbar. Er wird der Tochtergesellschaft DB Vertrieb GmbH zugeordnet. Das Unternehmen gibt an, dass der Chatbot in ständiger Weiterentwicklung ist. Über einen Avatar, dessen Name „Kai“ ist, trägt der Chatbot die Markenattribute des Unternehmens weiter. Sein Aufgabengebiet liegt in der Beantwortung von Fragen zu Produkten, Angeboten und Dienstleistungen der Deutsche Bahn AG. Kai antwortet auf Kundenanfragen, indem er diese nach Schlüsselwörtern gruppiert. Der Fokus liegt auf der unterstützenden Navigation im Onlinebereich. Die Lösung von vertragsspezifischen Kundenanfragen ist durch den Chatbot nicht möglich. Diese Belange werden über die Live-Chat Funktion mit einem Servicemitarbeiter, per Mail oder Telefon abgeschlossen (Deutsche Bahn AG, o.D.). Der Chatbot wird von dem deutschen Unternehmen e-bot7 entwickelt. E-bot7 ist darauf spezialisiert, KI in den bestehenden Kundenservice eines Unternehmens zu integrieren (e-bot7, o.D.). Die Speicherdauer von Kundenanfragen beträgt bis zu sieben Tage. Sie dient der Optimierung der Chatfunktionen durch selbstlernende Prozesse (Deutsche Bahn AG, o.D.). Der Chatbot Kai ist für den Nutzer jederzeit verfügbar. Im täglichen Zeitrahmen von 7 bis 20 Uhr wird der Chatbot unter Umständen von menschlichen Servicemitarbeitern unterstützt. Der Kunde erkennt an einem Bild neben dem Nachrichtendialog, ob er mit dem Chatbot (Kai) oder einem Servicemitarbeiter (Agent) kommuniziert. Die Interaktion findet ausschließlich in deutscher Sprache statt (Deutsche Bahn AG, o.D.).

### *Status quo des Kundenservice der SNCF*

Die Société nationale des chemins de fer français (SNCF) bedient diverse Kanäle, über welche Kunden den Kontakt zum Unternehmen aufnehmen können. Über eine Telefonrufnummer erfragt der Kunde Informationen zur aktuellen Verkehrssituation sowie zu Fahrplänen und ändert oder storniert eine Fahrt. Der Telefonservice der SNCF ist für den Kunden kostenlos und täglich von 7 Uhr bis 22 Uhr erreichbar. Über ein Schlüsselwort wird das Anliegen des Kunden automatisiert in eine Kategorie eingeteilt. Anschließend erfolgt eine Weiterleitung des Kunden zu einem verfügbaren Servicemitarbeiter. Auf der Webseite wird dem Kunden eine Übersicht mit den Wartezeiten der Hotline je nach Wochentag und Uhrzeit präsentiert. Es ist zusätzlich möglich, eine Beratung von einem regionalen Kundenberater zu erhalten. Weitere Rufnummern stehen bei Notfällen und internationalen Anliegen zur Verfügung. Einen zusätzlichen Kommunikationskanal stellen die sozialen Medien dar. Über Facebook, Twitter, Instagram, LinkedIn und YouTube können Kunden ihre Meinung äußern oder eine Frage stellen (SNCF, o.D.).

### *Chatbot OUIbot*

Der Chatbot der SNCF namens OUIbot unterstützt die Nutzer bei der Recherche nach Zugfahrplänen. Es ist möglich, über ihn einen Fahrkartenkauf der SNCF abzuschließen. SNCF kündigt an, dass es in der Zukunft möglich sein wird, die eigenen E-Tickets über den OUIbot abzurufen. Darüber hinaus annonciert das Unternehmen die Funktion zur Begleitung des Nutzers bis zum Bahnsteig. OUIbot wurde von der Tochtergesellschaft OUI.sncf entwickelt. Der Kunde erreicht den Chatbot über diverse Kommunikationsplattformen. Neben der Unternehmenswebseite ist der Chatbot über Facebook Messenger, WhatsApp und über Sprachbefehl mit Amazon Echo oder Google Home erreichbar. Zur Kommunikation mit dem OUIbot ist ein Benutzerkonto bei der SNCF obligatorisch. Der Chatbot kommuniziert mit dem Kunden ausschließlich in französischer Sprache (SNCF, o.D.).

### *Gegenüberstellung der charakteristischen Merkmale*

Der Chatbot der Deutsche Bahn AG trägt den männlichen Vornamen Kai (Deutsche Bahn, o.D.). Vereinzelt wird der Name auch von weiblichen Personen getragen, weshalb der Name Kai für beide Geschlechter Verwendung findet. Von dem Unternehmen SNCF wurde gleichermaßen kein Geschlecht des Chatbots festgelegt. Der Name OUIbot setzt sich aus dem französischen Begriff Oui (=Ja) und Bot, abgeleitet von dem Begriff Chatbot, beziehungsweise Robot zusammen. Der Name wird somit von den Kunden mit der Tochtergesellschaft oui.sncf assoziiert. Beide Chatbots vermitteln eine frohsinnige Ausstrahlung. Der Chatbot Kai beginnt stets die Konversation mit dem Kunden, indem er einen standardisierten Text versendet. In dieser Nachricht stellt er sich als virtueller Assistent in Ausbildung vor und erklärt seine Funktionen. Des

Weiteren bittet er den Kunden darum, seine Fragen möglichst kurz und prägnant zu verfassen (Deutsche Bahn, o.D.). In Bezug auf die Begrüßung des Kunden weist der OUIbot Parallelen zu Kai dem Chatbot auf. Der OUIbot stellt sich dem Kunden ebenfalls vor. Außerdem fordert er den Kunden auf, ihm das Reiseziel sowie die Reisedaten zu nennen. Der OUIbot verdeutlicht in der Konversation, dass er sich um das weitere Vorgehen kümmert, sobald er die gewünschten Informationen zu der Reise von dem Kunden erhält. Zusätzlich zeigt er die Möglichkeit auf, über die mobile Anwendung der SNCF oder der Unternehmenswebseite an Informationen zu der gebuchten Zugverbindung zu gelangen. Der OUIbot hält mit seinem einheitlichen, orangefarbenen Erscheinungsbild das Corporate Design des Unternehmens SNCF ein (SNCF, o.D.). Im Fall der Deutsche Bahn AG bestehen die Konversationsfelder und das Profilbild aus weiß und blauen Farbtönen. Diese stehen im Kontrast zu den Farben der Unternehmenswebseite, die in den Farben rot und weiß gehalten ist (Deutsche Bahn, o.D.).

### *Gegenüberstellung der Funktionalitäten*

Zur Gegenüberstellung der Performanz beider Chatbots werden zwei Szenarien einer Kundenanfrage simuliert. Die erste Kundenanfrage beinhaltet die Ermittlung von Informationen zu einem bestimmten Bahnhof. Diese Anfrage ist einer FAQ-Funktion zuzuordnen. Daraufhin erfolgt eine detaillierte Frage in Bezug auf die Fahrstuhlverfügbarkeit an einem genannten Bahnhof, der von den Verkehrsmitteln der jeweiligen Bahngesellschaft befahren wird. Der Chatbot Kai der Deutsche Bahn AG reagiert auf die Anfrage mit der Zusendung eines Links. Der erhaltene Link leitet den Kunden auf eine Internetseite der Deutsche Bahn AG weiter. Um die gewünschten Informationen zu erhalten, ist es notwendig, einen Bahnhof in das Suchfeld einzugeben. Anschließend werden die verfügbaren Ausstattungsmerkmale des ausgewählten Bahnhofs aufgezeigt. Es wird zusätzlich zu dem Anliegen die konkrete Frage gestellt, ob der Hauptbahnhof der Stadt Karlsruhe mit Fahrstühlen ausgestattet sei. Für diese Antwort wird der Chatbot von einem Servicemitarbeiter abgelöst. Dies erkennt der Kunde an dem beschrifteten Bild neben dem Konversationsfenster. Bei der Übernahme durch einen menschlichen Mitarbeiter erscheint die Bezeichnung „Agent“ an der Stelle des Namens des Chatbots „Kai“. Das simulierte Anliegen des Kunden wird von dem Chatbot Kai mit Unterstützung eines Servicemitarbeiters der Deutsche Bahn AG gelöst. Das beschriebene Szenario wird gleichermaßen bei dem Chatbot OUIbot der SNCF angewandt. Bei der Frage zu Informationen eines bestimmten Bahnhofs wird der Kunde auf die Unternehmensseite und den SNCF Assistent der mobilen Anwendung verwiesen. In Folge wird eine detaillierte Frage nach der Verfügbarkeit von Fahrstühlen am Bahnhof Straßburg gestellt. In diesem Fall gesteht der OUIbot, dass diese Frage seine Kompetenzen überschreitet. Folglich verweist er auf die Online-Support Seite des Unternehmens. Im zweiten

Szenario wird eine Anfrage zu einer vorgegebenen Reiseroute simuliert. Für die Chatbots der Deutsche Bahn AG und der SNCF erfolgt die Anfrage zu einer identischen Reiseroute von Karlsruhe nach Paris. Zusätzlich wird der Wunsch geäußert, die Reise noch am selben Abend antreten zu wollen. Der Chatbot Kai der Deutsche Bahn AG bittet als Reaktion auf die Anfrage darum, etwas Geduld bezüglich der Antwortgeschwindigkeit entgegenzubringen. Nach kurzer Zeit greift ein menschlicher Servicemitarbeiter ein. Es wird angemerkt, dass in der Konversation keine Auskunft über persönliche Preis- und Reiseauskünfte gegeben werden kann. Der Kunde wird über vier alternative Optionen zur Reiseplanauskunft aufgeklärt. Diese sind die mobile Anwendung der Deutsche Bahn AG, eine Verkaufsstelle sowie die Unternehmenswebseite. Die vierte Option stellt einen Anruf unter der aufgelisteten Servicehotline dar. Der Chatbot OUIbot der SNCF hingegen präsentiert dem Kunden eine Auswahl an möglichen Reisezeiten mit Preisangaben. Der Chatbot berücksichtigt dabei die bereits erworbenen Ermäßigungskarten des Kunden, wodurch eine personalisierte Preisanpassung erfolgt. Diese Information ist dem Chatbot durch die Anmeldung des Kunden in seinem Benutzerkonto bekannt. Der Besitz eines Kundenkontos ist für die Nutzung des Chatbots der SNCF obligatorisch. Weiterhin ist es möglich, die Buchung des Tickets mit dem OUIbot fortzuführen. Der Kunde wählt im Dialog eine Reiseverbindung aus. Anschließend fragt der Chatbot, ob die Angabe des Reisenden mit dem Konversationspartner übereinstimmt oder ob die Buchung für eine dritte Person erfolgt. Mit einem weiteren Klick bestätigt der Kunde die Reisedaten, den Preis sowie die allgemeinen Verkaufs- und Nutzungsbedingungen der SNCF. Außerdem stehen ihm die Optionen zur Verfügung, den Kaufvorgang abzubrechen oder eine Rückfahrt hinzuzufügen. Die Buchung ist damit beendet und wird dem Kunden angezeigt. Zusammenfassend lässt sich sagen, dass beide Chatbots in Bezug auf den ersten Gesamteindruck ähnlich erscheinen. Bei genauerem Betrachten der Funktionen, sind jedoch einige Unterschiede feststellbar. In Bezug auf FAQ-Funktionen ist der Chatbot der Deutsche Bahn AG nutzerfreundlicher gestaltet im Vergleich zu dem Chatbot der SNCF. Dies ist auf die Angabe der Querverweise im Dialog mit dem Chatbot Kai zurückzuführen, wodurch der Kunde auf die manuelle Eingabe der empfohlenen Informationsseite verzichten kann. Demgegenüber steht die Funktion zu individuellen Reiseverbindungen, bei dem der OUIbot im Vergleich zu dem Chatbot Kai eine hohe Performanz vorweist. Während der Chatbot Kai der Deutsche Bahn AG derzeit keine Funktion zur Anzeige der passenden Verbindungen für den Kunden auf Abruf bereitstellt, ist diese Funktion bei dem Chatbot der SNCF bereits vorhanden. Der OUIbot ermöglicht eine personalisierte Beratung des Kunden. Dies erfolgt mithilfe der Kundendaten, die im Benutzerkonto hinterlegt sind. Damit einhergehend ist die Begleitung des Kunden entlang der Customer Journey, da der Kunde mit dem

Chatbot der SNCF die verschiedenen Phasen von der Anfrage bis hin zum Kauf der Reiseverbindung durchläuft und von ihm in diese Richtung gelenkt wird. Dies verspricht eine einfache Nutzbarkeit, wodurch meist auch eine verbesserte Digital Customer Experience vorzuweisen ist. Die Voraussetzung einer vorherigen Registrierung als Kunde der SNCF, um mit dem OUIbot kommunizieren zu können, stellt jedoch auch einen Nachteil dar. Potenzielle Kunden profitieren nicht von der Kommunikation mit dem Chatbot. Dadurch ist die Generierung und Anziehung von Neukunden durch den Chatbot der SNCF nicht möglich (SNCF, o.D.); (Deutsche Bahn, o.D.).

## QUANTITATIVE FORSCHUNG

### Quantitatives Forschungsdesign

Charakteristisch für das quantitative Verfahren ist die Erfassung und Weiterverarbeitung von verbalen Informationen in Zahlenformate. Dieser Vorgang wird Quantifizierung genannt (Reinders, 2011, S. 121). Zur Sicherstellung der Vergleichbarkeit erfolgt eine Auswertung in Form von standardisierten, anonymen Fragebögen. Diese werden über eine Teilgesamtheit von Merkmalsträgern durch eine schriftliche, internetgestützte Befragung beantwortet (Eckstein, 2014, S. 41). Diese Methode wird gewählt, da eine Umfrage die Befragung subjektiver Bewertungen und Einstellungen ermöglicht (Möhring und Schlütz, 2019, S.5). Dies begünstigt die Beantwortung der Forschungsfrage, die sich auf die Wahrnehmung von Chatbots im Kundenservice bezieht. Sinnvoll erscheint die Wahl der Online-Umfrage zudem, da sich das Thema der vorliegenden Arbeit auf Chatbots bezieht, die sich auf digitalen Touchpoints bewegen.

### Erhebungsinstrument und Aufbau der Befragung

Das Erhebungsinstrument der quantitativen Forschung ist der Fragebogen einer standardisierten Befragung. Der entwickelte Fragebogen besteht aus vier thematischen Blöcken. Der erste Teil beginnt mit den Erfahrungen des Teilnehmers mit Chatbots und seiner Wahrnehmung zu den bisherigen Interaktionen. Im zweiten Teil werden die Anforderungen an einen zufriedenstellenden Kundenservice abgefragt. Daraufhin folgt die Analyse der Kundenzufriedenheit in Bezugnahme auf den Kundenservice des Unternehmens. Im vierten Themenbereich werden demographische Daten zu der Person sowie ihre digitale Affinität festgehalten. Das zweisprachige Projekt enthält inhaltlich identische Fragen mit dem Unterschied, dass sich die Fragen in Teil drei auf unterschiedliche Unternehmen in der Mobilitätsbranche beziehen. Die deutschsprachigen Probanden erhalten in diesem Teil Fragen zu der Deutsche Bahn AG, während den französischsprachigen Probanden Fragen zu der SNCF gestellt werden. Die Abfolge der Fragen erfolgt nach der Ordnung von allgemeinen Fragen zu spezifischen Fragen des Mobilitätsunternehmens. Die Fragen sind leicht



verständlich formuliert, um eine erleichterte Beantwortung des Fragebogens sicherzustellen.

### Festlegung & Struktur der Stichprobe

Zur Untersuchung der Problematik und der anschließenden Prüfung der Hypothesen ist eine klare Definition der erforschten Personengruppen erforderlich. Da in dem vorliegenden Fall eine Vollerhebung der Grundgesamtheit aufgrund der Größe nicht möglich ist, wird eine Stichprobe mit einer Teilmenge der betreffenden Gruppe erhoben. Das Ziel der Stichprobe liegt in der Generalisierung auf die Grundgesamtheit. Es ist von einem Repräsentativitätsschluss auszugehen (Raithel 2008, S. 54f.). Die Grundgesamtheit stellen dabei Mitbürger dar, die in Frankreich oder in Deutschland leben und das Internet in ihrem Alltag nutzen. Da die quantitative Analyse mithilfe einer internetgestützten Befragung durchgeführt wird, ist diese Voraussetzung bei den beantworteten Fragebögen stets gegeben. An der Umfrage nahmen insgesamt 274 Personen teil. Davon beantworteten 201 Personen den Fragebogen bis zum Schluss. Es besteht eine etwa gleichmäßige Verteilung von deutschsprachigen und französischsprachigen Teilnehmern der Umfrage, sodass die französischsprachige Variante eine Teilnehmeranzahl von 98 Personen enthält und die deutschsprachige Variante eine Anzahl von 103 Teilnehmern vorweist. Die Beendigungsquote in Höhe von 73,4 Prozent wird auf das innovative Thema zurückgeführt. Im weiteren Verlauf wird die Stichprobe anhand der demographischen Daten beschrieben. Diesbezüglich werden die Teilnehmer am Ende des Fragebogens nach ihrem Geschlecht und Alter befragt.

### Pretest

Vor dem Beginn der Feldphase wird das Erhebungsinstrument einem Pretest unterzogen. Im Rahmen der vorliegenden quantitativen Untersuchung wird der Pretest an einer begrenzten Anzahl von zehn ausgewählten Merkmalsträgern durchgeführt. Bei der Auswahl der Personen wird auf ein ausgewogenes Verhältnis zwischen der französischsprachigen und deutschsprachigen Beantwortung der Fragebögen geachtet (Raithel, 2008, S. 63).

### Fragebogenkonstruktion

Die Fragestruktur ist ausschließlich von geschlossenen Fragen mit einer vorgegebenen Antwortkategorie geprägt. Zur Gewährleistung der Vergleichbarkeit und der Auswertungsobjektivität der mehrsprachigen Fragebogenausführungen wird von halboffenen und geschlossenen Fragen abgesehen. Es handelt sich somit um eine standardisierte Befragung. Ferner wird die Problemstellung im nächsten Schritt im Rahmen der qualitativen Analyse mithilfe von offenen Fragen exploriert. In der quantitativen Untersuchung stellen Ratingskalen zur Abfrage von Bewertungen, Intensitäten und Häufigkeit die dominierende Antwortkategorie dar. Die Antwortkategorien sind präzise, erschöpfend und

disjunkt konstruiert. Es wird eine eindimensionale, gerade Skalierung festgelegt, um eine Positionierung der befragten Person zu erwirken. Unter dem Einsatz der geraden Skalierung wird die Tendenz zur Mittelkategorie unterbunden (Raithel, 2008, S.67-71).

### Deskriptive und induktive Auswertung &

#### Überprüfung der Hypothesen

##### *Demographische Daten*

Unter den 103 deutschsprachigen Teilnehmer gehören 26 Prozent dem männlichen Geschlecht an, während 74 Prozent dem weiblichen Geschlecht zuzuordnen sind. Unter den 98 französischsprachigen Teilnehmern ist die Mehrheit mit 67 Prozent weiblichen Geschlechts während 33 Prozent dem männlichen Geschlecht angehören.

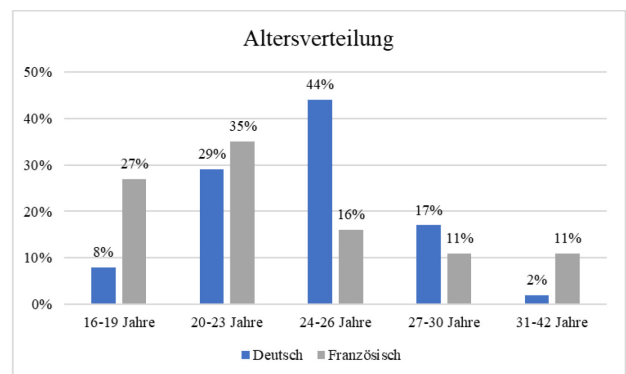


Abbildung 1: Altersverteilung

Der Median der Altersverteilung deutschsprachiger Teilnehmer beträgt 24 Jahre, während der Median der französischsprachigen Teilnehmer bei 22,5 Jahren liegt. Das Alter stellt eine Variable dar, die eine hohe Zahl an Ausprägungen vorweist. Aus Gründen der Übersichtlichkeit werden die Werte in eine neue Variable in fünf Altersklassen umcodiert. Es lässt sich feststellen, dass die Mehrheit der Befragten jünger als 26 Jahre alt ist. Das heißt, dass überwiegend Teilnehmer vorliegen, die der Generation Y (Personen, die im Zeitraum der frühen 1980er bis zu den späten 1990er Jahren geboren wurden) und der Generation Z (Nachfolgegeneration der Generation Y) angehören. Die Zusammensetzung der Stichprobe ist für die Auswertung des Fragebogens vorteilhaft, da die Teilnehmer eine relevante Gruppe für die Analyse der Wahrnehmung von Chatbots repräsentieren.

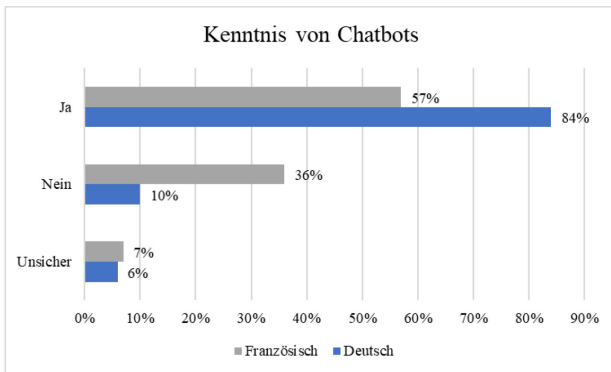


Abbildung 2: Kenntnis von Chatbots

Aus Abbildung 2 wird ersichtlich, dass die Mehrheit der französischsprachigen Teilnehmer (57 Prozent) Chatbots kennen, während der Anteil der deutschsprachigen Teilnehmer, die Chatbots kennen, bei 84 Prozent liegt. Für diejenigen Teilnehmer, die angeben unsicher zu sein oder Chatbots nicht zu kennen, wird im nächsten Schritt eine Definition von Chatbots eingeblendet, um in der Lage zu sein, die folgenden Fragen in Bezug auf Chatbots zu beantworten. Zur Überprüfung von  $H_{01}$  wird ein Chi-Quadrat-Test durchgeführt, sodass eine Beurteilung des Zusammenhangs zweier nominalskalierten Merkmale in Bezug auf ihre Abweichung der statistischen Unabhängigkeit erfolgen kann (Kähler, 2011, S. 105). Ein Mehrfelder-Chi-Quadrat Test nach Pearson ergibt höchst signifikante Unterschiede bei den Häufigkeiten der Kenntnis von Chatbots zwischen Deutschland und Frankreich ( $\chi^2 = 20,57$ ;  $p < 0,001$ ; Cramér-V = 0,32). Der Cramér-V-Koeffizient ist identisch zu dem Wert des Phi-Koeffizienten ( $\phi$ ) und hat mit  $V = 0,32$  eine mittlere Effektstärke. Demzufolge kann die Nullhypothese  $H_{01}$  verworfen werden.

### Chatbots

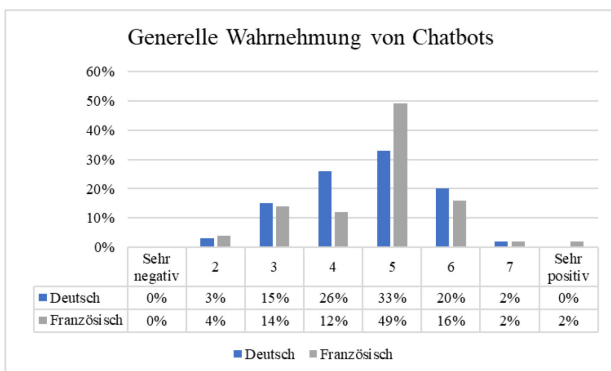


Abbildung 3: Generelle Wahrnehmung von Chatbots

In Bezug auf die Nutzung von Chatbots zur Interaktion mit Unternehmen werden die Teilnehmer aufgefordert, ihre generelle Wahrnehmung in einer achtstufigen Skala mit Endpunktbenennung einzuordnen, wie aus Abbildung 3 ersichtlich wird. Es ist festzustellen, dass keiner der Teilnehmer die Nutzung von Chatbots zur Kommunikation mit Unternehmen als sehr negativ einstufte. Außerdem wird die Nutzung lediglich von zwei Prozent der französischsprachigen Teilnehmer und von

keinem der deutschsprachigen Teilnehmer als sehr positiv eingestuft. Die Mehrheit der Antworten befindet sich bei den Teilnehmern aus Frankreich sowie aus Deutschland auf der Stufe 5 der Skala. Somit empfinden 33 Prozent der Deutschen und 40 Prozent der Franzosen die Nutzung von Chatbots zur Interaktion mit Unternehmen als eher positiv. Das arithmetische Mittel liegt bei den deutschsprachigen Ergebnissen bei 4,57 und bei den französischsprachigen bei 4,73. Aufgrund des hinreichend großen Stichprobenumfangs erfolgt ein parametrischer Test zur Überprüfung von  $H_{02}$ . Der t-Test für unabhängige Stichproben setzt voraus, dass die Merkmale ein intervallskaliertes Skalenniveau vorweisen, was in diesem Fall zutrifft. Aus den 103 deutschsprachigen und den 98 französischsprachigen Teilnehmern ergibt der Mittelwertvergleich mit Hilfe eines t-Tests für unabhängige Stichproben keinen signifikanten Mittelwertunterschied ( $p = 0,97$ ) bei einer Irrtumswahrscheinlichkeit von 5 Prozent (Kähler, 2011, S.409).  $H_{02}$  kann somit nicht verworfen werden.

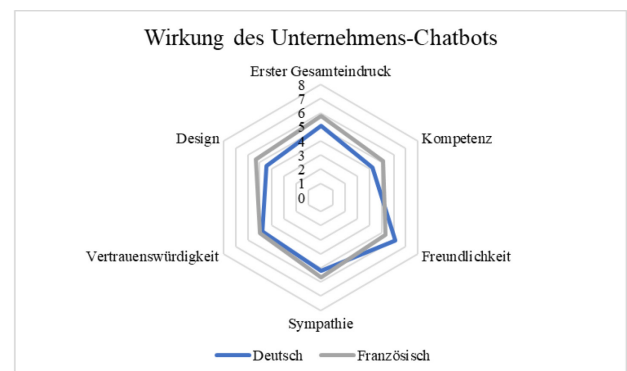


Abbildung 4: Wirkung des Unternehmens-Chatbots

Abbildung 4 stellt die Wirkung des Unternehmens-Chatbots auf die Studienteilnehmer dar. Hierbei wird den Teilnehmern im Voraus eine Bildschirmaufnahme des Unternehmens-Chatbots gezeigt. Dabei erhalten die französischsprachigen Teilnehmer eine Bildschirmaufnahme des OUIbots der französischen SNCF und die deutschsprachigen Teilnehmer eine Aufnahme des Service-Chatbots Kai der Deutsche Bahn AG. Die Befragten werden dazu aufgefordert, auf einer achtstufigen Skala (1 = sehr negativ, 8 = sehr positiv) die oben dargestellten Kriterien zu bewerten. Zur besseren Übersichtlichkeit werden die Ergebnisse der jeweiligen Kriterien in Abbildung 4 in Form des arithmetischen Mittels dargestellt. Es ist festzustellen, dass die Einordnungen der französischsprachigen Teilnehmer im Mittel um den Wert 5 anzutreffen sind. Dies stellt eine eher positive Einstellung zu dem Unternehmens-Chatbot in den Bereichen Kompetenz, Freundlichkeit, Sympathie, Vertrauenswürdigkeit, Design sowie dem ersten Gesamteindruck dar. Die Ergebnisse der deutschsprachigen Fragebögen liegen im Mittel bei den Werten vier und fünf. Lediglich die wahrgenommene Freundlichkeit des deutschen Unternehmens-Chatbots weicht mit dem durchschnittlichen Wert von einer 6,15 („ziemlich positiv“) ab. Allgemein lässt sich feststellen, dass der Chatbot der französischen Bahngesellschaft

(5,76) im Gesamteindruck positiver wahrgenommen wird als der Chatbot der Deutsche Bahn AG (5,1). Zur Prüfung von  $H_{03}$  wird ein Mittelwertvergleich mit Hilfe eines t-Tests für unabhängige Stichproben herangezogen, welcher einen sehr signifikanten Mittelwertunterschied von 0,61 ( $t = -3,28$ ,  $p < 0,01$ ) zwischen dem ersten Gesamteindruck des Unternehmens-Chatbots und den beiden Ländern ergibt. Die Nullhypothese  $H_{03}$  kann somit verworfen werden.

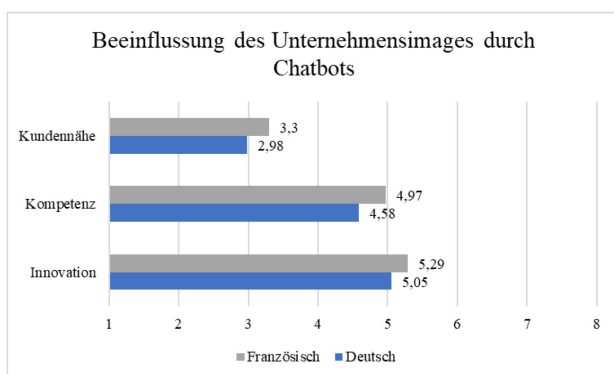


Abbildung 5: Beeinflussung des Unternehmensimages durch Chatbots

Abbildung 5 stellt die Beeinflussung des Unternehmensimages unter den Aspekten der wahrgenommenen Kundennähe, Kompetenz und Innovation durch den Einsatz von Chatbots dar. Auf der achtstufigen Likert-Skala beantwortet der Teilnehmer sein Empfinden in den Bereichen „Trifft gar nicht zu“ (1) bis „Trifft sehr zu“ (8). Es ist hervorzuheben, dass jeweils die Mittelwerte der Ergebnisse in der Abbildung 5 dargestellt sind, um eine bessere Vergleichbarkeit zu gewährleisten. Das Markenattribut Kundennähe wird von den französischsprachigen Teilnehmern im Mittel mit einem Wert von 3,3 angegeben, welcher zwischen der Antwortmöglichkeit „Trifft ziemlich zu“ und „Trifft eher nicht zu“ vorzufinden ist. Das Ergebnis der deutschsprachigen Befragten liegt mit einem Mittelwert von 2,98 leicht darunter. Es lässt sich feststellen, dass die Antworten der französischsprachigen Teilnehmer im Durchschnitt eine höhere Zustimmung zu dem Zusammenhang des Unternehmensimages mit den drei vorgegebenen Markenattributen aussprechen. Der Aussage, dass Unternehmen die Chatbots einsetzen, kompetent wirken, stimmen die französischsprachigen Teilnehmer im Durchschnitt eher zu (4,97), während der Durchschnitt der deutschsprachigen Teilnehmer dieser Ausführung im Vergleich etwas weniger zustimmt. Dieses Bild ergibt sich ebenfalls bei dem Markenattribut Innovation. In diesem Zusammenhang wird  $H_{04}$  überprüft. Der Mittelwertvergleich zu der wahrgenommenen Kompetenz von Unternehmen, die Chatbots einsetzen, ergibt einen signifikanten Mittelwertunterschied von 0,39 ( $t = 1,99$ ;  $p < 0,05$ ). Dieser Wert setzt sich aus der Differenz der Mittelwerte von Deutschland (4,58) und Frankreich (4,97) zusammen. Der Mittelwertunterschied, den die beiden Gruppen aufweisen, beträgt 0,282

Standardabweichungen auf der jeweiligen Skala in Relation zu der Streuung der Variablen. Die Stärke des Ergebnisses (0,282) hat somit einen kleinen Effekt ( $|d| \approx 0,2$ ). Die wahrgenommene Innovation im Zusammenhang mit Unternehmen, die Chatbots einsetzen, unterscheidet sich nicht signifikant zwischen den beiden Ländern. ( $p = 0,27$ ). Ebenso ergibt der Mittelwertvergleich bei der wahrgenommenen Kundennähe zu Unternehmen mit Chatbot-Einsatz keinen signifikanten Unterschied zwischen Deutschland und Frankreich ( $p = 0,17$ ). In Bezug auf die Wahrnehmung der Kompetenz von Unternehmen, die Chatbots einsetzen, wird die Nullhypothese  $H_{04}$  verworfen.

### Digitale Affinität

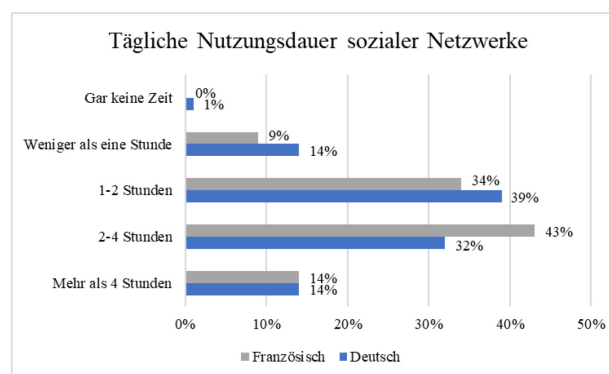


Abbildung 6: Tägliche Nutzungsdauer sozialer Netzwerke

Abbildung 6 gibt die tägliche Nutzungsdauer sozialer Netzwerke wieder. Die Mehrheit der deutschsprachigen Teilnehmer (39 Prozent) nutzt soziale Netzwerke 1-2 Stunden täglich, wohingegen die Mehrheit der französischsprachigen Teilnehmer (43 Prozent) eine tägliche Nutzung von 2-4 Stunden angibt. Ein übereinstimmendes Ergebnis bietet die Antwortoption der Nutzung von sozialen Netzwerken von mehr als 4 Stunden täglich, welche in beiden Sprachvariationen einem Anteil von 14 Prozent entspricht. Weiterhin gibt lediglich ein Prozent der deutschsprachigen Teilnehmer an, gar keine Zeit auf sozialen Netzwerken zu verbringen, während keiner der französischsprachigen Teilnehmer vollkommen auf die Anwendung von sozialen Netzwerken verzichtet.

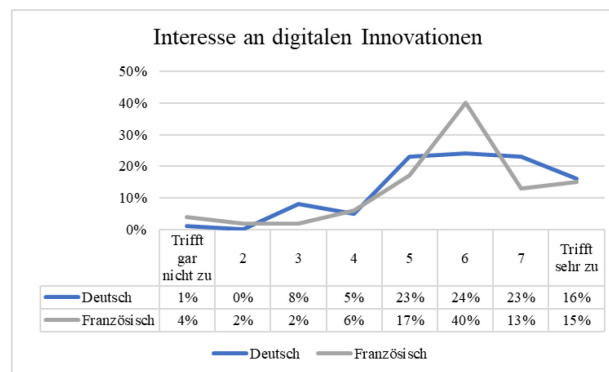


Abbildung 7: Interesse an digitalen Innovationen

Abbildung 7 stellt die Ergebnisse einer weiteren Aussage zu der digitalen Affinität der Teilnehmer auf einer achtstufigen Likert-Skala dar. Diese werden von den Teilnehmern in Bezug auf ihre Zustimmung eingeordnet. Es ist unverkennbar, dass die Mehrheit der französischsprachigen Teilnehmer mit einem Anteil von 40 Prozent der Aussage ziemlich zustimmt (6). Der Graph erreicht an dieser Stelle den Höhepunkt. Dahingegen ist die Mehrheit der Antworten der deutschsprachigen Teilnehmer mit einem Anteil von 24 Prozent derselben Antwortmöglichkeit nicht in gleicher Weise aussagekräftig, da die beiden benachbarten Antwortkategorien lediglich in geringem Maß abweichen. Im Bereich der ablehnenden Antwortkategorien (1-4) ist ein Anteil von 14 Prozent der deutschsprachigen Teilnehmer und ein abweichend verteilter Anteil von insgesamt 14 Prozent der französischsprachigen Teilnehmer vorzufinden. Das arithmetische Mittel liegt für Deutschland bei 5,93 und für Frankreich bei 5,8. Zur Überprüfung von  $H_{05}$  werden die digitale Affinität sowie die zuvor präsentierte Wahrnehmung von Chatbots im Kundenservice benötigt. Zur Messung der digitalen Affinität der Teilnehmer werden die Variablen der verbrachten Zeit auf Social Media und das Interesse an Innovationen herangezogen. Die Korrelation zwischen der generellen Wahrnehmung von Chatbots und der verbrachten Zeit auf Social Media ist signifikant ( $r = -0,14$ ;  $p < 0,05$ ). Die negative Korrelation hat ihren Ursprung aus der Reihenfolge der Antwortmöglichkeiten der Variable zu der verbrachten Zeit auf sozialen Medien. Diese sind absteigend sortiert. Mithilfe einer Umcodierung der Antwortmöglichkeiten zu einer aufsteigenden Reihenfolge erhält man schließlich eine positive Korrelation der beiden Variablen. Ebenso ergibt sich eine signifikante, positive Korrelation zwischen den Variablen der generellen Wahrnehmung von Chatbots und dem Interesse an Innovationen ( $r = 0,15$ ;  $p < 0,05$ ). Basierend auf den beiden signifikanten Korrelationen der Wahrnehmung von Chatbots mit der verbrachten Zeit auf Social Media und dem Interesse an Innovationen, die als Indikatoren der digitalen Affinität dienen, kann die Nullhypothese  $H_{05}$  verworfen werden.

### Kundenservice

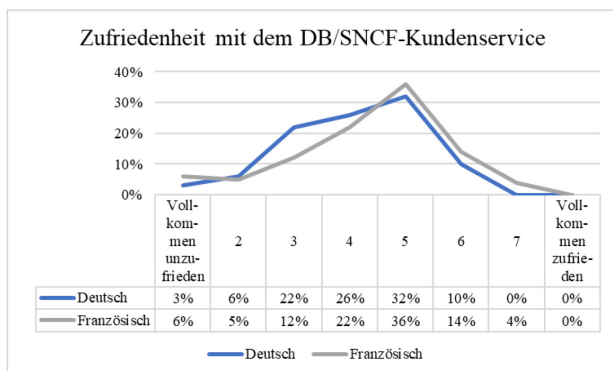


Abbildung 8: Zufriedenheit mit dem Kundenservice der DB respektive der SNCF

Abbildung 8 illustriert die Zufriedenheit mit dem Kundenservice des Unternehmens, zu dem die Teilnehmer befragt werden. Die französischsprachigen Teilnehmer beantworten hierbei die unternehmensspezifischen Fragen in Bezug auf die SNCF, während die deutschsprachigen Teilnehmer die Fragen zu der Deutsche Bahn AG beantworten. Die Mehrheit beider Gesellschaften ist mit dem Kundenservice des Unternehmens eher zufrieden, wobei 32 Prozent der deutschsprachigen und 36 Prozent der französischsprachigen Teilnehmer diese Meinung vertreten. Im Gegenzug ist keiner der Befragten vollkommen zufrieden mit dem Kundenservice, während drei Prozent der Deutschen und sechs Prozent der Franzosen vollkommen unzufrieden sind. Im Antwortbereich der Unzufriedenheit (1-4) lässt sich ein leicht erhöhter Anteil der deutschsprachigen Teilnehmer im Vergleich zu den französischsprachigen Teilnehmern feststellen. Im Umkehrschluss ist der Anteil der französischsprachigen Teilnehmer im Bereich der Zufriedenheit (5-8) etwas höher als der Anteil der deutschsprachigen Teilnehmer. In Bezug auf  $H_{06}$  ergibt sich eine höchst signifikante Korrelation zwischen den Variablen der Kundenzufriedenheit der Bahngesellschaften und des Gesamteindrucks des Unternehmens-Chatbots ( $r = 0,3$ ;  $p < 0,001$ ). Dies bedeutet, je zufriedener ein Kunde mit dem Kundenservice der jeweiligen Bahngesellschaft ist, desto positiver bewertet er den ersten Gesamteindruck des Unternehmens-Chatbots. Die Nullhypothese  $H_{06}$  kann somit verworfen werden.

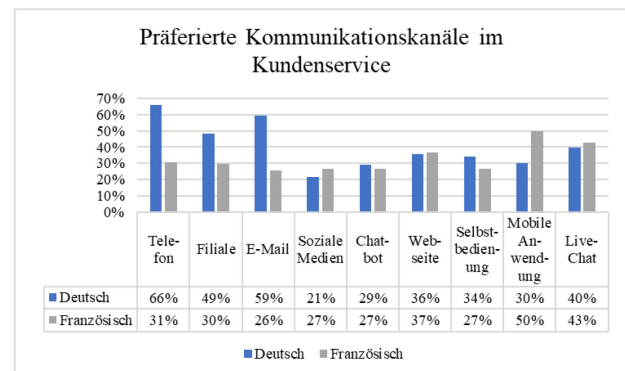


Abbildung 9: Präferierte Kommunikationskanäle im Kundenservice

Abbildung 9 präsentiert die Antworten der Befragten in Bezug auf die Frage, welche Kommunikationskanäle sie bei der Kontaktaufnahme mit dem Kundenservice der jeweiligen Bahngesellschaft bevorzugen. Hierbei ist eine Mehrfachauswahl der Antwortkategorien möglich, sodass die dichotomen Variablen im Vorfeld als Variablensets definiert werden, um die jeweiligen Anteile der Items berechnen zu können. Ein eklatanter Unterschied ist in der Präferenz des Telefons als Kommunikationskanal zwischen den beiden Kulturen festzustellen. Während 31 Prozent der französischsprachigen Teilnehmer das Telefon zur Kommunikation mit Unternehmen gerne nutzen, sind es in Deutschland 66 Prozent, die dieses Mittel bevorzugen.

Das Telefon stellt somit für die deutschsprachigen Teilnehmer das bevorzugte Kommunikationsmittel dar, um mit dem Kundenservice der Bahngesellschaft in Kontakt zu treten. Für die französischsprachigen Teilnehmer stellt die mobile Anwendung den präferierten Kommunikationskanal mit einem Anteil von 50 Prozent dar. Die Kommunikation mit einem Unternehmens-Chatbot können sich 29 Prozent der Deutschen und 27 Prozent der Franzosen vorstellen. Zur Überprüfung von  $H_{07}$  wird ein Mittelwertvergleich mit Hilfe eines t-Tests für unabhängige Stichproben durchgeführt, welcher einen höchst signifikanten Mittelwertunterschied von 0,35 ( $t = 5,34$ ;  $p < 0,001$ , Cohen's  $d = -10,65$ ) bei der Variable Telefon als Kommunikationskanal ergibt. Das bedeutet, dass deutsche Personen die Kontaktaufnahme zum Kundenservice eines Unternehmens mit dem Telefon eher präferieren als französische Personen. Darüber hinaus ergibt die Variable des E-Mail-Verkehrs als Kommunikationskanal einen höchst signifikanten Mittelwertunterschied beider Länder von 0,33 ( $t = 5,13$ ;  $p < 0,001$ ; Cohen's  $d = -10,71$ ). Diese Signifikanz zeigt, dass deutsche Kunden im Vergleich zu französischen Kunden eher dazu neigen, eine E-Mail an den Kundenservice zu schreiben. Darüber hinaus ergibt der Mittelwertvergleich im Rahmen des t-Tests für unabhängige Stichproben der Variable des persönlichen Kontaktes in einer Filiale einen sehr signifikanten Mittelwertunterschied von 0,19 ( $t = 2,8$ ;  $p < 0,01$ ; Cohen's  $d = -10,4$ ). Es ist daher festzustellen, dass deutsche Personen im Vergleich zu französischen Personen den persönlichen Kontakt in einer Filiale bevorzugen. Demgegenüber ergibt der Mittelwertvergleich der Variable der mobilen Anwendung als präferierter Kommunikationskanal einen signifikanten Mittelwertunterschied von 0,2 ( $t = -2,92$ ;  $p < 0,05$ ; Cohen's  $d = -10,36$ ). Das bedeutet, dass französische Personen im Vergleich zu deutschen Personen den Kommunikationskanal der mobilen Anwendung zur Kontaktaufnahme bevorzugen. Die Kommunikationskanäle Soziale Medien ( $p = 0,39$ ), Unternehmenswebseite ( $p = 0,91$ ), Selbstbedienung im Internet ( $p = 0,25$ ) und Live-Chat mit einem Servicemitarbeiter ( $p = 0,66$ ) ergeben keinen signifikanten Mittelwertunterschied im Ländervergleich. Gleichermaßen verhält es sich mit dem Kommunikationskanal des Unternehmens-Chatbots im Vergleich zwischen Deutschland und Frankreich. Auch dieser weist keinen signifikanten Mittelwertunterschied ( $p = 0,68$ ) auf. In Bezug auf die signifikanten Mittelwertunterschiede der drei Variablen Telefon, Filiale und mobile Anwendungen kann die Nullhypothese  $H_{07}$  verworfen werden.

### Interpretation der quantitativen Ergebnisse

Es lässt sich feststellen, dass es einen Unterschied in dem Kenntnisstand von Chatbots zwischen Deutschland und Frankreich gibt. Demnach kennen mehr deutsche als französische Personen Chatbots. Weiterhin wurden keine signifikanten Unterschiede in der generellen Wahrnehmung von Chatbots beobachtet. Beide Länder

stufen Chatbots im mittleren Bereich mit einer positiven Tendenz ein. Die Ergebnisse weisen einen exakt identischen Mittelwert auf, was in dieser Höhe der Stichprobengröße zwar unüblich ist, jedoch auf eine ausgeglichene Wahrnehmung beider Länder deutet. Weiterhin ist es ersichtlich, dass es Unterschiede in der Wahrnehmung des Unternehmens-Chatbots der jeweiligen Bahngesellschaft gibt. Der französische Chatbot OUIbot wurde positiver bewertet als der deutsche Chatbot Kai. Diese Differenz hängt jedoch nicht nur mit den kulturellen Unterschieden zusammen, sondern vielmehr mit dem unterschiedlichen Auftritt der beiden Chatbots. Der OUIbot ist in das Corporate Design der SNCF eingebettet, während der Chatbot Kai weniger Parallelen zu dem Corporate Design der Deutsche Bahn AG aufweist. Außerdem ist ein Unterschied in der Wahrnehmung der Kompetenz eines Unternehmens durch den Einsatz von Chatbots im Ländervergleich festzuhalten. Die Beeinflussung des Unternehmensimages durch Chatbots wirkt sich demzufolge positiv aus. Französische Personen nehmen diesen Einfluss im Mittel positiver wahr als deutsche Personen. Dies ist auf die geringere Unsicherheitsvermeidung der französischen Kultur nach den Kulturdimensionen von GLOBE zurückzuführen (Brodbeck 2016, S. 130-136). Weiterhin ist ein Zusammenhang der digitalen Affinität und der Wahrnehmung von Chatbots feststellbar. Bei der Prüfung dieser Hypothese handelt es sich primär um die Differenzen zwischen digital affinen Personen und weniger digital affinen Personen in Bezug auf ihre Wahrnehmung auf Chatbots. Da die Mittelwerte der digitalen Affinität beider Länder nicht stark voneinander abweichen, lässt sich diese Erkenntnis in beiden Fällen anwenden. Eine positivere Wahrnehmung von Chatbots kann durch die digitale Affinität verbessert werden, da die Hemmschwelle der Person dadurch geringer ist. Personen, die mit der digitalen Welt vertraut sind, befinden sich dementsprechend in der Lage, sich auf neue Technologien einzulassen. Darüber hinaus stellt ein häufiger Gebrauch von sozialen Netzwerken eine weitere digitale Kommunikationsmöglichkeit dar, von der Chatbots nicht weit entfernt sind, unter Anbetracht der Tatsache, dass diese bereits auf diversen Kommunikationsplattformen im Einsatz sind. Weiterhin bestätigt sich ein positiver Zusammenhang zwischen der Kundenzufriedenheit mit der Bahngesellschaft und dem Unternehmens-Chatbot. Das bedeutet, dass ein zufriedener Kunde eher dazu geneigt ist, den Unternehmens-Chatbot positiv wahrzunehmen. Auffallend ist die unterschiedliche Präferenz der Kommunikationskanäle der beiden Länder. Während deutsche Personen die traditionellen Kommunikationskanäle (Telefon, Filiale, E-Mail) bevorzugen, präferieren die französischen Personen mobilen Anwendungen zur Kontaktaufnahme mit dem Kundenservice der SNCF. Diese Vorliebe lässt sich ebenfalls auf die kulturelle Dimension der Unsicherheitsvermeidung zurückführen, in denen

Frankreich niedrigere Ergebnisse in der Praktik als Deutschland aufweist.

## QUALITATIVE FORSCHUNG

Mithilfe der qualitativen Forschung erfolgt die Weiterführung einer bereits abgeschlossenen, quantitativen Forschung. Die weiterführende Forschung hat zum Zweck, die Plausibilität inferenzstatistischer Zusammenhänge zu reflektieren und zu erhärten (Schulz, 1977 zitiert nach Mayring, 2008, S. 20f.). Als Erhebungsinstrument wird die qualitative Befragung gewählt. Die Interviewerin und die jeweils Befragten sitzen sich persönlich gegenüber. Die Fragen der Interviewerin werden mündlich mitgeteilt. Außerdem wird das Gespräch zur späteren Auswertung aufgezeichnet (Möhring und Schlütz, 2019, S. 126). Da die Wichtigkeit von konkreten Aussagen im Vordergrund steht, wird das Leitfadenterview zur mündlichen Befragung gewählt. Dies erfolgt in einem halbstrukturierten Rahmen (Mayer, 2013, S. 37). Die qualitative Auswertung der Ergebnisse erfolgt gemäß der Globalauswertung nach Legewie (1994, S. 177-182). Diese hat zum Ziel, eine übersichtliche Auswertung zu ermöglichen.

### Interpretation der qualitativen Ergebnisse

Die Meinungen beider Befragten unterscheiden sich in Bezug auf die Kenntnis von Chatbots. Die Befragte 1 sieht die Hintergründe der unterschiedlichen Kenntnisstände im Alter, während die Befragte 2 der Meinung ist, dass die Unterschiede zwischen Deutschland und Frankreich auf den kulturellen Differenzen beruhen. Die Befragte 1 sieht die visuelle Erscheinung des Unternehmens-Chatbots als zentrales Element der Wahrnehmung, während die Befragte 2 die Freundlichkeit dessen als besonders relevant betrachtet. Beide Befragten sind der Meinung, dass Chatbots das Unternehmensimage der Bahngesellschaften beeinflussen könnten. Dies könnte sich im Fall eines positiven Outcomes verbessern, jedoch sei auch eine Verschlechterung des Unternehmensimages vorstellbar, sofern der Chatbot nicht die Erwartungen des Kunden erfüllen würde. Der Einfluss der digitalen Affinität einer Person auf die Wahrnehmung von Chatbots stößt bei beiden Seiten auf unterschiedliche Auffassungen. Die Befragte 1 stimmt dieser Hypothese nicht zu, da sie der Meinung ist, dass Chatbots keine Notwendigkeit für digital affine Nutzer darstellen. Die Befragte 2 betont, dass sie sich diesen Zusammenhang sehr gut vorstellen könne und schätzt eine gleichmäßige Verteilung dieser Hypothese in beiden Ländern ein. Eine Gemeinsamkeit ergibt sich aus der Auffassung, dass weniger digitale Personen die persönliche Interaktion mit Kundendienstmitarbeitern bevorzugen. Im nächsten Schritt erfolgt eine Stellungnahme der beiden Interviewtexte in Bezug auf den literarischen Kontext. Die Befragte 1 begründet eine unterschiedliche Wahrnehmung von Chatbots im Kundenservice mit den Altersunterschieden. Die Ergebnisse der quantitativen

Umfrage ergaben jedoch keinen signifikanten Zusammenhang zwischen dem Alter und der Wahrnehmung der Befragten. Es ist jedoch hervorzuheben, dass die Mehrheit der Befragten der Generation Z und der Generation Y angehören. In diesem Rahmen wäre eine größere Altersverteilung der älteren Generationen wünschenswert, um repräsentative Schlüsse daraus ziehen zu können. Allerdings sollten die kulturellen Unterschiede weiterhin Beachtung finden. Gemäß der Kulturdimensionen nach GLOBE existieren unterschiedliche Ausprägungen der Unsicherheitsvermeidung und der Humanorientierung. Diese Faktoren beeinflussen somit die Wahrnehmung von Chatbots. Ein relevanter Punkt, den die Befragte anspricht, ist das Potenzial von Chatbots zur Seitennavigation der Unternehmenswebseite. Somit wird der Kunde zu seinem Ziel begleitet, wodurch die Customer Experience optimiert werden kann. Die Befragte 1 sieht den Onlineservice als komplementäre Dienstleistung zu dem traditionellen Kundenservice. Dabei ist hervorzuheben, dass Unternehmen sich den steigenden Kundenanforderungen stetig anpassen. Aus diesem Grund ist eine Omnipräsenz der beiden Bahngesellschaften unerlässlich. Die Befragte 2 sieht den Ursprung der Unterschiede in den kulturellen Dimensionen. Der hohe Stellenwert des persönlichen Kontaktes ist durch die höhere Praktik der Humanorientierung Frankreichs zu begründen. Die Präferenz der mobilen Anwendung auf französischer Seite bildet hierbei jedoch ein Paradox, da auf die menschliche Interaktion gezielt verzichtet wird. Die niedrigere Praktik der Unsicherheitsvermeidung in Frankreich zeigt, dass die französische gegenüber der deutschen Kultur neuen Technologien offener gegenübersteht. Das begründet die positivere Wahrnehmung des Unternehmens-Chatbots der SNCF. Diese positive Wahrnehmung ist jedoch von weiteren Kriterien abhängig. So ist in diesem Kontext die umfassendere Funktionalität des französischen Chatbots zu beachten. Ferner gehört die Farbgebung des Unternehmens-Chatbots, der bei der SNCF in das Corporate Design des Unternehmens eingebettet ist dazu. Entsprechend besteht bei dem Unternehmens-Chatbot der Deutsche Bahn AG Optimierungsbedarf (Persönliche Kommunikation, 2020).

### FAZIT

Die Ergebnisse der quantitativen und qualitativen Forschung ergaben einige Zusammenhänge wie auch Unterschiede der Wahrnehmung zwischen Deutschland und Frankreich. Lediglich in der globalen Wahrnehmung von Chatbots wurden keine signifikanten Differenzen festgestellt. Bei der Betrachtung einzelner Aspekte zeigen jedoch die Unterschiede in der Wahrnehmung des Unternehmens-Chatbots der beiden Bahngesellschaften SNCF und Deutsche Bahn auf, dass der französische Chatbot in seinem Gesamteindruck positiver bewertet wird. Die Hintergründe dieser Differenz wurden im qualitativen Teil erörtert. Dabei entstand die Annahme, dass das Design des Chatbots sowie die aufgebrachte

Freundlichkeit gegenüber dem Kunden ausschlaggebende Kriterien sind. Weiterhin ergab die Studie, dass der Einsatz von Chatbots im Kundenservice Einfluss auf das Unternehmensimage nimmt. Hierbei nehmen französische Kunden das Unternehmen als kompetenter wahr als deutsche Kunden. Dieses Ergebnis ist in den Kontext der kulturellen Dimensionen einzuordnen, die einige Unterschiede zwischen den Kulturen aufzeigen. Die Dimension der Unsicherheitsvermeidung ist demnach in der deutschen Kultur stärker ausgeprägt. Somit stehen Mitglieder der französischen Gesellschaft neuen Technologien offener gegenüber. Diese Annahme bestätigt sich gleichermaßen in der Wahl der präferierten Kommunikationskanäle zur Interaktion mit einem Unternehmen. Die deutschen Kunden wählen vorzugsweise traditionelle Kommunikationskanäle, zu denen das Telefon, die E-Mail sowie der persönliche Kontakt in der Filiale zählen. Im Gegenzug präferieren französische Kunden mobile Anwendungen zur Kontaktaufnahme. Diese Tendenz ist auf die geringere Unsicherheitsvermeidung zurückzuführen. Einen Widerspruch ergibt jedoch die stärkere Ausprägung der Humanorientierung der französischen Kultur, da eine Präferenz des persönlichen Kontaktes annehmbar wäre. Diese Präferenz bestätigt sich in den qualitativen Interviews. Ferner wurden einige länderübergreifende Zusammenhänge festgestellt. Je digital affiner ein Kunde ist, desto positiver ist seine Wahrnehmung gegenüber Chatbots. Dieser Effekt bestätigt sich durch die quantitative Forschung und durch die Zustimmung einer Befragten aus den geführten Leitfadeninterviews. Einen weiteren Zusammenhang ergibt die Kundenzufriedenheit mit der jeweiligen Bahngesellschaft und der Bewertung des Unternehmens-Chatbots. Diese Annahme ist in beiden Ländern anwendbar und bestätigt die Relevanz einer allgemeinen Kundenzufriedenheit. Für eine optimierte Wahrnehmung von Chatbots der Mobilitätsbranche in Deutschland und in Frankreich sind einige Aspekte zu beachten. Um den Einsatz von Chatbots im Kundenservice in Deutschland und Frankreich zu optimieren, empfiehlt es sich, die kulturellen Unterschiede zu berücksichtigen.

## ZUKÜNFTIGE FORSCHUNG

Diese Studie hat einen ersten Forschungsbeitrag zur Wahrnehmung von Chatbots im Kundenservice in der Mobilitätsbranche im Vergleich zwischen Deutschland und Frankreich geleistet. Der Aspekt des Einflusses von Chatbots auf die Digital Customer Experience könnte einen bedeutenden Ansatz für die zukünftige Forschung bieten. Für die weitere Forschung eignet sich ein Testexperiment, in dem der Proband den Kundenservice über einen traditionellen Kommunikationskanal kontaktiert und anschließend über einen Chatbot. Daraufhin werden Fragen zur Einschätzung des Käuferlebnisses gestellt. Mithilfe dieses Testprojekts ist eine direkte Gegenüberstellung der Kommunikationskanäle in Bezugnahme auf den Einfluss der Customer Experience möglich. Weiterhin behandelt die vorliegende Arbeit ausschließlich die Wahrnehmung

von textbasierten Chatbots im Kundenservice. Im Rahmen einer weiteren Forschung wäre es interessant, den Fokus um die Perzeption von sprachgesteuerten Assistenten im Kundenservice zu erweitern. Somit würde durch die Sinneserweiterung des Hörens ein weiterer Faktor in die Beobachtung einfließen.

## LITERATUR

- Bendel, Oliver (2019), "Chatbot," (Zugriff am Dezember 28, 2019), [verfügbar unter <https://wirtschaftslexikon.gabler.de/definition/chatbot-54248/version-277297>].
- Borowski, Craig (2015), "What a Great Digital Customer Experience Actually Looks Like," (Zugriff am Dezember 28, 2019), [verfügbar unter <https://hbr.org/2015/11/what-a-great-digital-customer-experience-actually-looks-like>].
- Brodbeck, Felix C. (2016), *Internationale Führung. Das GLOBE-Brevier in der Praxis. Die Wirtschaftspsychologie*. Berlin, Heidelberg: Springer.
- Eckstein, Peter P. (2014), *Statistik für Wirtschaftswissenschaftler. Eine realdatenbasierte Einführung mit SPSS*. Wiesbaden: Springer Fachmedien Wiesbaden GmbH.
- Frohne, Greta (2020), *Erfolgsfaktoren von Online Customer Experience. Die Gestaltung von internetbasierten Kundenerlebnissen in der Konsumgüterbranche. BestMasters*. Wiesbaden: Springer Fachmedien Wiesbaden; Springer Gabler.
- Gouthier, Matthias, Hrsg. (2016), *Kundenbindung durch kosteneffiziente Service Excellence. Strategien - Konzepte - Best-Practices: Konferenzband der 5. Excellence-in-Service-Konferenz EXIS. Reihe Dienstleistungsmanagement, Dienstleistungsmarketing, Band 1*. Baden-Baden: Nomos.
- Grand View Research (2017), "Chatbot Market Size To Reach \$1.25 Billion By 2025 | CAGR: 24.3%," (Zugriff am Januar 22, 2020), [verfügbar unter <https://www.grandviewresearch.com/press-release/global-chatbot-market>].
- Hartmann, Florian G. und Daniel Lois (2015), *Hypothesen Testen*. Wiesbaden: Springer Fachmedien Wiesbaden.
- Holland, Heinrich (2019), "Customer Experience Management," (Zugriff am Dezember 28, 2019), [verfügbar unter <https://wirtschaftslexikon.gabler.de/definition/customer-experience-management-54478/version-277507>].
- Hussy, Walter, Margrit Schreier, und Gerald Echterhoff (2013), *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor. Springer-Lehrbuch*. Berlin, Heidelberg: Springer.
- Kelle, Udo (2019), "Mixed Methods," in *Handbuch Methoden der empirischen Sozialforschung*, Nina Baur und Jörg Blasius, Hrsg. Wiesbaden: Springer Fachmedien Wiesbaden, 159–170.
- Kuckartz, Udo (2014), *Mixed Methods: Springer Fachmedien Wiesbaden*.
- Legewie, Heiner (1994), "Globalauswertung von Dokumenten," in *Texte verstehen. Konzepte, Methoden, Werkzeuge. Schriften zur Informationswissenschaft*, Nr. 14, Andreas Boehm, Hrsg. Konstanz: Univ.-Verl. Konstanz, 177–182.
- Mayer, Horst O. (2013), *Interview und schriftliche Befragung*. München: Oldenbourg.
- Mayring, Philipp (2008), *Qualitative Inhaltsanalyse. Grundlagen und Techniken. Beltz Pädagogik*. Weinheim: Beltz.

- Möhring, Wiebke und Daniela Schlütz (2019), *Die Befragung in der Medien- und Kommunikationswissenschaft. Eine praxisorientierte Einführung. Studienbücher zur Kommunikations- und Medienwissenschaft*. Wiesbaden: Springer Fachmedien Wiesbaden; Springer VS.
- Raithel, Jürgen (2008), *Quantitative Forschung. Ein Praxiskurs. Lehrbuch*. Wiesbaden: VS Verl. für Sozialwissenschaften.
- Reinders, Heinz (2011), *Empirische Bildungsforschung. Lehrbuch*. Wiesbaden: VS-Verl.
- Robra-Bissantz, Susanne und Christoph Lattemann, Hrsg. (2019), *Digital customer experience. Mit digitalen Diensten Kunden gewinnen und halten. Edition HMD*. Wiesbaden, Germany: Springer Vieweg.
- Rühle, Alexander, Lavinia Hoesch, und Marko Petersohn (2019), "Herausforderungen in der Mensch-Maschine-Interaktion durch den Einsatz von Bots," in *Innovativer Einsatz digitaler Medien im Marketing*, Winnen, Hrsg. Wiesbaden: Springer Fachmedien Wiesbaden, 47–62.
- Stanoevska-Slabeva, Katarina (2018), "Conversational Interfaces — die Benutzerschnittstelle der Zukunft?" *Wirtschaftsinformatik & Management*, 10 (6), 26–37.
- Kähler, Wolf-Michael (2011), *Statistische Datenanalyse. OnlinePlus*. [Erscheinungsort nicht ermittelbar]: Vieweg+Teubner.
- Stucki, Toni (2018), *Chatbots gestalten mit Praxisbeispielen der schweizerischen Post. Hmd best paper award 2018*. [S.l.]: Morgan Kaufmann.
- Töpfer, Armin und Andreas Mann (2008), "Kundenzufriedenheit als Basis für Unternehmenserfolg. Welche Bedeutung kommt der Zufriedenheit von Kunden für den Markterfolg des Unternehmens zu?" in *Handbuch Kundenmanagement. Anforderungen, Prozesse, Zufriedenheit, Bindung und Wert von Kunden*, Armin Töpfer, Hrsg. Berlin, Heidelberg: Springer, 38–79.

## KONTAKT

Daniela Mema  
 Hochschule Pforzheim  
 Tiefenbronner Straße 65  
 75175 Pforzheim  
 Daniela.Mema@outlook.de

Prof. Dr. Véronique Goehlich  
 Hochschule Pforzheim  
 Tiefenbronner Straße 65  
 75175 Pforzheim  
 Veronique.Goehlich@hs-pforzheim.de

Prof. Dr. Frank Morelli  
 Hochschule Pforzheim  
 Tiefenbronner Straße 65  
 75175 Pforzheim  
 Frank.Morelli@hs-pforzheim.de



# ITIL4 und DevOps – Gegensätze, Komplemente oder ein lösbarer Konflikt?

Carsten Dorrhauer

Fachbereich III

Hochschule Ludwigshafen  
Ernst-Boehe-Straße 4  
D-67059 Ludwigshafen  
E-Mail: carsten.dorrhauer@hwg-lu.de

Haio Röckle

Fachbereich III

Hochschule Ludwigshafen  
Ernst-Boehe-Straße 4  
D-67059 Ludwigshafen  
E-Mail: haio.roeckle@hwg-lu.de

## ABSTRACT

Die Entwicklung der Softwaretechnik einerseits und die Verbreitung von ITSM-Prozessen andererseits erfordern immer dringlicher die Verzahnung von DevOps und ITIL. Unterdessen ist aber deren Kompatibilität noch nicht systematisch untersucht. Erschwert wird die Gestaltung von Prozessen, die den Anforderungen von ITIL und DevOps Genüge tragen, durch die Verbreitung heterogener Landschaften sowohl technischer als auch organisatorischer Art. Aus technischer Sicht werden Cloud-Dienste mit On-Premise-Diensten kombiniert, aus organisatorischer Sicht wird dem Kunden ein IT-Service angeboten, der von mehreren Dienstleistern jeweils in Teilen angeboten wird. Hier soll der Frage nachgegangen werden, ob und unter welchen Bedingungen beide Konzepte vereinbar sind. Besonderes Augenmerk liegt dabei auf ITIL4, das seit 2020 vorliegt und das an einigen Stellen explizit auf DevOps Bezug nimmt.

## SCHLÜSSELWÖRTER

IT-Service-Management, ITIL, Software Engineering, DevOps

### Einleitung

Seit langem herrscht weitgehender branchenweiter Konsens über die Notwendigkeit professioneller Prozesse für das IT-Service-Management. Die mit großem Abstand weiteste Verbreitung haben die Prozesse und Praktiken der IT Infrastructure Library (ITIL) gefunden, die ihren Ursprung vor rund 30 Jahren in britischen Regierungsorganisationen haben. Neben vielen anderen Aspekten regeln sie die Bereitstellung von Soft- und Hardwarereleases sowie von Patches zur Fehlerbehebung, deren Versionierung, Vorkehrungen zu ihrer Wiederherstellung im Störungs- und Katastrophenfall, Abläufe des Roll-Outs und ähnliches.

Seit einigen Jahren nun gewinnt das Konzept DevOps eine inzwischen ähnlich weite Verbreitung in der Softwareindustrie. Es zielt auf die Integration von Entwicklungs- und Betriebsprozessen. Letztere sind aber in vielen Organisationen die Domäne von ITIL.

Damit drängt sich die Frage auf, inwiefern diese Ansätze sich ausschließen, sich ergänzen oder sich kombinieren lassen. Im Folgenden sollen deshalb die aktuellen Standardwerke zu beiden Konzepten miteinander in Beziehung gesetzt werden. Im ersten Falle ist das die quasi standardgebende Axelos, die die Rechte an ITIL hält. Im zweiten Fall sind es die Publikationen der meistbeachteten DevOps-Protagonisten, allen voran Patrick Debois und Gene Kim.

### Methodik

Zunächst sollen in aller Kürze die beiden Konzepte ITIL und DevOps mit ihrem jeweiligen Fokus vorgestellt

werden. Eine Harmonisierung der Prozesse des Software Engineering und des IT-Service-Management steht aktuell vor der zusätzlichen Herausforderung, auf heterogene und virtualisierte Umgebungen Rücksicht nehmen zu müssen, auf die deshalb in Folge eingegangen wird. Anschließend wird der Frage nachgegangen, welche ITSM-Prozesse dieser Harmonisierung bedürfen. Schließlich wird die Vereinbarkeit der beiden Konzepte im Allgemeinen untersucht, um dann darzustellen, wie die aktuellen ITIL4-Publikationen im Besonderen diese Thematik explizit und implizit aufgreifen.

### IT-Service-Management mit ITIL

ITIL ist eine Sammlung von sogenannten Best Practices, die auf einen Leitfaden für das IT-Service-Management zurückgehen, der ab 1989 für die britische Central Computer and Telecommunications Agency (CCTA) entwickelt wurde. (Vgl. Böttcher 2008, S. 1). Inzwischen legt jährlich eine sechsstellige Zahl von Menschen die Prüfungen ab, um ITIL-Zertifikate zu erwerben. (Vgl. Corless 2017). Diese Verbreitung ist für Praktiken und Prozesse in der IT-Industrie einzigartig und nur mit der Marktdurchdringung führender Softwareprodukte zu vergleichen.

### Software Engineering mit DevOps

2009 organisierte Patrick Debois die ersten DevOps-Days in Gent. Grundlegende Idee waren und sind *gemeinsame* Ziele, Abläufe und Werkzeuge (Goals, Processes, Tools) für Entwicklungsprojekte *und* Systembetrieb. Damit soll die klassische und bis heute häufige Dichotomie von IT-Organisationen abgemildert und am besten schließlich überwunden werden, weil sie häufig zu Konflikten, unübersichtlichen Abläufen und unklaren

Zuständigkeiten führt und letztlich die Produktivität der Gesamtorganisation behindert.

DevOps ist im Grunde unabhängig vom Vorgehensmodell, ergibt aber aus Sicht des Projektmanagements besonders zusammen mit agilen Methoden Sinn. Aus technischer Sicht wird DevOps i.d.R. mit Methoden der Continuous Integration kombiniert. Die Kombination geeigneter Werkzeuge ermöglicht das mindestens tägliche Zusammenführen der Entwicklungsstände aller beteiligten Entwickler.

### **Heterogene und virtualisierte Umgebungen**

ITIL adressiert u.a. Betriebsprozesse. DevOps adressiert u.a. Deployment-Prozesse. Beide Aufgaben werden immer komplexer angesichts heterogener Systemlandschaften und angesichts der Verbreitung virtualisierter Lösungen, insbesondere in der Cloud.

So muss z.B. ein Incident-Management-Prozess nach ITIL zur Behebung von Betriebsstörungen durch multiple Teams mit unterschiedlichen Kompetenzfeldern für Störungsursachen auf unterschiedlichen Plattformen komplex gestaltet werden.

Ein automatisiertes Deployment muss z.B. neben der eigentlich auszuliefernden Software auch Konfigurationsdateien berücksichtigen. (Infrastructure as code). Beide Konzepte müssen auch auf diese Herausforderung gemeinsam reagieren, um erfolgreich zu sein. Die Konfiguration der VM, auf denen die mit DevOps auszuliefernde Anwendung läuft, insbesondere in der Cloud, unterliegt genau wie die Software selbst dem Change und Release Management nach ITIL.

### **Betroffene ITSM-Prozesse**

ITIL beschreibt eine Vielzahl von Practices, die die Vorgehensweise eines IT-Service-Providers abdecken sollen. Viele davon haben keine oder keine wesentlichen Berührungspunkte zum Software Engineering und damit zu DevOps. Bei ITIL stehen der Servicelebenszyklus und die kontinuierliche Serviceverbesserung im Mittelpunkt, bei DevOps der Softwarelebenszyklus und die kontinuierliche Verbesserung von Software. Aus Sicht von ITIL ist eine Software ein Configuration Item (CI). Da viele Services primär mittels Software-CIs erstellt werden, gibt es in diesem Umfeld Überschneidungen. Zu den Practices, die davon betroffen sind, gehören sicherlich vor allem *Change Control*, *Release Management* und *Deployment Management*. Sie befassen sich mit der Kontrolle von Änderungen an der Infrastruktur, mit dem Management von Releases, die i.d.R. mehrere solcher Änderungen umfassen sowie mit den technischen Aspekten der Auslieferung von neuer oder geänderter Software, Hardware und Dokumenten in eine Produktiv- oder auch Testumgebung.

Weitere Practices sind eher mittelbar betroffen, so das *Availability Management* sowie das *Capacity and Performance Management*. Für beide ist Software eine von mehreren CI bei der Erstellung eines Service, die seine Verfügbarkeit und Kapazität mitbestimmen. In einer DevOps-Organisation benötigen beide Practices Zielvorgaben und Informationen aus DevOps-Teams.

Die operative Arbeit mit dem Benutzer am *Service Desk* beim *Incident Management* nach ITIL ist offensichtlich mit DevOps kompatibel. Sie verändert sich höchstens insofern, dass im Rahmen des 2nd- oder 3rd-Level-Supports Entwickler an der Bearbeitung von Störungsmeldungen beteiligt werden und damit Teil der Prozesskette werden - falls das nicht ohnehin längst der Fall sein sollte. Analog lässt sich für den Prozess *Problem Management* argumentieren, der nach Fehlerursachen sucht und weitere Störungen vermeiden soll.

Zu den wesentlichen Voraussetzungen einer erfolgreichen DevOps-Organisation gehört die Erfassung und Auswertung von Produktiv-Telemetriedaten, um „immer genug Telemetriedaten (zu) haben, ... dass sich ... Services im Produktivumfeld korrekt verhalten. Und wenn Probleme auftreten, ... schnell herausfinden (zu) können, was falsch läuft, und ausreichend Informationen haben...“ (Kim et al. 2017, S. 190) Die Herausforderung ist dabei weniger eine technische als eine organisatorische, namentlich die Überwindung von Daten-Silos und die Definition von Zuständigkeiten und Prozessen für die Auswertung der Logs und die Erzeugung von Metriken.

Exakt diese Voraussetzung von DevOps schafft ITIL mit dem V3-Prozess *Event Management*, der in der ITIL4-Practice *Monitoring and Event Management* aufgeht.

### **Kompatibilität zwischen DevOps und ITIL**

Bereits in der Einleitung zu ihrem DevOps-Standardwerk zählen die wohl wichtigsten DevOps-Protagonisten als einen von ihnen kurz zu widerlegenden Mythos über DevOps den Satz auf: „DevOps ist nicht kompatibel zu ITIL.“ (Kim et al. 2017, S. XVI) Sie argumentieren noch auf Basis von ITIL V3, dass DevOps kompatibel zu ITIL-Prozessen gestaltet werden könne, aber einer weitgehenden Automatisierung der ITIL-Prozesse bedürfe (vgl. ebenda, S. XVII) Umgekehrt diene die mit DevOps einhergehende höhere Deploymentgeschwindigkeit und –frequenz der Aktualität der Konfigurationsdatenbank CMDB und der Softwarebibliothek DML nach ITIL. (vgl. ebenda)

Eine der wichtigsten und ersten Maßnahmen zur organisatorischen Implementierung von DevOps-Prinzipien ist die Definition von Zuständigkeiten, Rollen und Kommunikationsbeziehungen an der Schnittstelle zwischen Entwicklung und Betrieb. (vgl. Kim et al. 2017, S. 91ff.) Das kann im Einzelfall unterschiedlich aussehen. Denkbar sind

- der Einbezug von Operatoren in die Entwicklungsteams, z.B. als Teilnehmer des Daily Scrum
- Self-Service-Angebote
- sogenannte Ops-Liaisons, also eigene Rollen an der Schnittstelle zwischen Entwicklung und Betrieb (vgl. ebenda, S. 93)

Da ITIL zwar vielfältige Zuständigkeiten, Rollen und Kommunikationsbeziehungen vorschlägt, darüber hinausgehende zusätzliche Zuständigkeiten, Rollen und

Kommunikationsbeziehungen aber nur in seltenen Ausnahmen ausschließt, überrascht es nicht, dass dieser Schritt in Richtig DevOps in keiner Weise mit ITIL im Konflikt steht.

Im Gegenteil fordert ITIL genau den Übergang von Kooperation zu Kollaboration zwischen den Teams in einer IT-Organisation und meint damit die Zusammenarbeit an gemeinsam vereinbarten Zielen. (vgl. Axelos 2020, S. 22) Dies deckt sich mit der Forderung nach gemeinsamen Zielsetzungen von Entwicklung und Systembetrieb bei DevOps. (vgl. Kim et al. 2017, S. 73)

Es bleibt die Frage zu klären, ob DevOps womöglich mit operativen ITIL-Prozessen, allen voran Change Management (ab ITIL4: Change Control), in Konflikt steht. Change Control hält die Balance zwischen Flexibilität und Zuverlässigkeit. Dabei gilt aber die Prämisse: "This assessment... should not introduce an unnecessary delay." (Axelos 2019, S. 119)

Insbesondere das mit DevOps einhergehende automatisierte Deployment könnte schließlich von restriktiven Service-Management-Prozessen ausgebremst werden. Automatisiertes Deployment vieler kleinerer Änderungen stellt offensichtlich andere Anforderungen an die Operations-Prozesse als ein großer Releasewechsel. Vereinfacht oder erschwert also automatisiertes Deployment die SM-Prozesse? Vereinfachen oder erschweren SM-Prozesse automatisiertes Deployment? Die Antwort findet sich in der Change-Kategorisierung nach ITIL. Nicht alle Änderungen müssen vom zuständigen Gremium, dem Change Advisory Board (CAB), einzeln autorisiert werden. *Standard Changes* können vorab genehmigt werden; freilich müssen sie dennoch dokumentiert werden. (Vgl. Axelos 2019, S. 119) Das typische Beispiel dafür ist die Installation von Standardsoftware auf Clients, die bereits auf anderen Clients installiert ist. Die Worst-Case-Betrachtung liefert das Bild eines einzigen ausgefallenen Arbeitsplatzes – ein Risiko, das in Kauf genommen werden kann. Diese Kategorie kann natürlich auch auf serverseitige Changes angewandt werden. Entscheidend ist die Risikobewertung, die ihrerseits abhängt vom „Vertrauen in Kontrollelemente wie automatisierte Tests und das proaktive Produktivitäts-Monitoring.“ (Kim et al. 2017, S. 325) Wichtig für ein reibungsloses, weitgehend automatisiertes Deployment ist, dass das CAB vorab nachvollziehbare Kriterien für Standard-Changes definiert, auf die die Entwickler sich verlassen können. Verschmelzen Entwicklungs- und Betriebsorganisation, so wird das CAB, das über diese Vorab-Genehmigung entscheidet, u.a. mit Entwicklern besetzt sein. Und automatisiertes Deployment erfordert ohnehin penible Versionierung und Qualitätskontrolle. Wenn man so will, legen damit die Entwicklungsprojekte aus eigener Motivation eine Disziplin diesbezüglich an den Tag, die das Service Management schon lange einfordert. Kim et al kommentieren einen Fall aus der Praxis wie folgt: „Lieg darin nicht eine Ironie? Dev beschwert sich gern darüber, dass Ops sich scheut, den Code zu deployen. Aber in diesem Fall haben die Entwickler die Macht, ihren eigenen Code zu deployen, und trauen sich genauso

wenig.“ (Vgl. Kim et al 2017, S 221) DevOps-Teams, die Verantwortung für ihre Deployments tragen, scheinen genau zu der Vorsicht zu tendieren, die seinerzeit im Operations erst zur Etablierung von formalen Change-Prozessen geführt hat.

Die Kategorisierung als *Standard Change* kann viele, aber nicht alle Deployments vereinfachen. Im Lebenszyklus jeder Software gibt es Updates, die mit einem gewissen Risiko behaftet sind und die konsequenterweise vom CAB zu genehmigen sind, ITIL bezeichnet sie als *Normal Changes*. Unterstellt man ein grundsätzlich vertrauensvolles Verhältnis zwischen Entwicklung und Betrieb im allgemeinen – ohne das DevOps sowieso zum Scheitern verurteilt wäre – und innerhalb des CAB im besondern, so steht und fällt die Qualität und Geschwindigkeit seiner Entscheidungen mit der Verfügbarkeit ausreichender Informationen zu den Änderungsanforderungen (RfC). Kim et al sind dazu optimistisch: „Wir können ziemlich sicher das Erstellen vollständiger und genauer RfCs automatisieren und das Ticket mit den Details zur Änderung versehen.“ (Kim et al 2017, S. 328). Genau dies sieht ITIL ausdrücklich vor: „...organizations ... often automate most steps of the change control process.“ (Axelos 2019, S. 119)

Stehen also Peer-Reviews aus agilen Modellen im Widerspruch zum formalen Change Control durch ein dafür vorgesehenes Gremium wie das CAB? Nur, wenn sie es in allen Fällen ersetzen sollten und aus überschneidungsfreien Personengruppen bestünden, die gegensätzliche Interessen verfolgen würden. Dies ist aber nicht notwendigerweise der Fall.

Teams, die nach DevOps arbeiten, brauchen schnelle Deployments. ITIL kann diese ermöglichen, wenn die Beteiligten es wollen. Ein Change Manager, der einen RfC grundsätzlich ablehnt, weil eine Information fehlt, anstatt mit dem Entwicklungsteam Rückfrage zu halten, wird ein DevOps-Team ausbremsen. So kann der falsche Eindruck entstehen, ITSM-Prozesse seien ein Hindernis für DevOps.

Hat ein geplantes Deployment den Change-Control-Prozess durchlaufen, so können die Installationen stattfinden. ITIL fordert einen gesicherten Speicherort, der Kopien aller verwendeten Softwaresysteme in den verschiedenen Versionen enthält, die Definitive Media Library (DML). Bei Bedarf kann aus der DML eine bestimmte Software jederzeit zu einem gewünschten Stand wiederhergestellt werden, unabhängig von den beteiligten Personen. Realisiert wird die DML für Individualsoftware meist als privater Cloudspeicher, für Standardsoftware gegebenenfalls als physischer Ablageort optischer oder anderer Datenträger. Kann nun ein automatisiertes und häufiges Deployment mit der DML in Übereinstimmung gebracht werden? Offensichtlich ist das eher eine organisatorische als eine technische Herausforderung. Erstens muss in der DevOps-Toolchain ein schreibender Zugriff auf die DML integriert sein. Auf diese Weise lässt sich im Störfall, z.B. nach dem Austausch einer Festplatte oder eines Servers, zweifelsfrei exakt die Softwareversion wiederherstellen, die in der Produktivumgebung zuletzt betrie-

ben wurde – auch an Feiertagen, mitten in der Nacht oder lange nachdem die verantwortlichen Entwickler das Unternehmen verlassen haben. Alternativ ist auch ein Pull-Mechanismus denkbar, bei dem die DevOps-Tools nicht schreiben, sondern den Schreibvorgang anstoßen. Zweitens müssen die ITIL-Prozesse für Change Control und Deployment Management natürlich genau diesen Schreibzugriff gestatten. Auch hierbei steht und fällt der Erfolg mit dem Zusammenwirken der Zuständigen für ITSM und DevOps, der personellen Überschneidung der Teams sowie angemessen zugeschnittenen Zielvereinbarungen und Verantwortlichkeiten.

### Lösungen bei ITIL4

ITIL versteht sich als Konzept für IT-Dienstleister (IT Service Provider, ITSP). Indem es die Perspektive von IT-Produkten auf IT-Dienstleistungen lenkt, fasst es die Entwicklung der ihnen zugrunde liegenden Produkte mit deren Inbetriebnahme, Betrieb und dem Benutzersupport zusammen. Zum Design des Service gehört mehr als das Design einer Anwendungssoftware. Es umfasst u.a. die Vorbereitung der SLA und der Lieferantenbeziehungen sowie die Planung der Kapazität, der Sicherheit und der Verfügbarkeit nicht nur der Hardware, sondern z.B. auch des Personals. (vgl. Axelos 2020, S. 70) Zur Inbetriebnahme des Service gehört u.a., die Rollen und Prozesse des Service Desk aufzusetzen, wobei dem Early-Life Support (ELS) besondere Bedeutung zukommt. (vgl. ebenda, S. 73f.) Die Trennung von Dev und Ops ist bereits dem Kerngedanken von ITSM fremd.

Die Protagonisten von ITIL (vgl. Axelos 2020, S. 9) und DevOps (vgl. Kim et al 2017, S. 83) sind sich einig, dass statisches technisches Expertentum, das traditionell weit in der IT verbreitet ist, den Anforderungen moderner IT-Organisationen nicht mehr genügt. Organisationen benötigen Mitarbeiter mit breiten und flexiblen Kompetenzprofilen. Was DevOps-Teams benötigen, sind Teammitglieder, die beide Felder, Dev und Ops, abdecken können. Dies deckt sich mit dem „pi-shaped individual“ nach ITIL. (vgl. Axelos 2020, S. 11) Dort wird auch beschrieben, wie Organisationen die Rahmenbedingungen setzen können, um die Entwicklung solcher Kompetenzprofile zu fördern. (vgl. ebenda). Ähnlich, aber ausführlicher und mit starkem Fokus auf das Lernen aus Fehlern, wird bei DevOps argumentiert. (Vgl. Kim et al 2017, S. 265ff.)

Für einen ITSP, der nach DevOps arbeitet, kann die Bildung marktorientierter Teams sinnvoll sein, die „funktionsübergreifend und unabhängig“ (Kim et al. 2017, S. 78) gestaltet werden. Sie ermöglichen flexible Kollaboration zwischen Entwicklung und Systembetrieb mit einem begrenzten Fokus. Die Herausforderung, die sich daraus ergibt, liegt in der Binnenorganisation dieser multifunktionalen Einheiten. ITIL4 schlägt als Lösung Swarming vor. Dabei bearbeiten Experten verschiedener Felder, hier also Dev und Ops, solange gemeinsam eine Aufgabe, bis sich herausstellt, wer am geeignetsten ist, sie weiterzuverfolgen. (Vgl. Axelos 2020, S. 94)

Wohl am offenkundigsten tritt der Einfluss von DevOps auf die ITIL4-Practices beim V3-Prozess *Release and Deployment Management* zutage. DevOps braucht schnelle und weitgehend automatisierte Lösungen dafür. Detaillierter werden in ITIL4 Werkzeuge und Methoden beschrieben, mit deren Hilfe dieses Ziel erreicht werden kann. (Axelos 2020, S.34ff) Dass in ITIL4 die Practices *Release Management* und *Deployment Management* separat aufgeführt und beschrieben werden, ist sicherlich nicht nur dem Umstand geschuldet, dass ITIL4 endlich zwischen Service Management Practices und Technical Management Practices unterscheidet und beide entsprechend einsortiert. Vielmehr erleichtert diese Unterscheidung die Verwendung von ITIL in DevOps-Umgebungen, wenn zwischen einem Release einerseits sowie einem oder mehreren Deployments andererseits sauber unterschieden wird. Beide können nunmehr zeitlich und inhaltlich unabhängig voneinander betrachtet, erstellt, validiert und genehmigt werden. Ein Deployment muss nicht unmittelbar den Benutzern verfügbar gemacht werden, erst Recht nicht der Gesamtheit der Benutzer. Ein Vorgehen der Art Test - Pilotierung - allgemeine Verfügbarkeit ohne erneutes Deployment war in ITIL3 schwer abbildbar, ggf. hätten Dummy-Releases erstellt werden müssen, um dem Prozess Genüge zu tun. Um dies abbilden zu können, berücksichtigt die ITIL4-Practice *Release Management* nun insbesondere Blue/Green-Releases in zwei gespiegelten Produktivumgebungen (vgl. Axelos 2019, S. 135) sowie Feature Flags zur kontrollierten Freischaltung einzelner Features für einzelne Nutzer oder Nutzergruppen. (vgl. ebenda, S. 136) Völlig unproblematisch ist aus Sicht von DevOps die Zuständigkeit für Release-Entscheidungen bei ITIL: “The tools ... may be the responsibility of a dedicated person, but decisions about the release can be made by the development team.“ (Ebenda)

Deployment Management ist in ITIL4 zu einer eigenen Practice im neu geschaffenen Abschnitt *Technical Management Practices* befördert worden. Es erlaubt nahe-liegenderweise inkrementelle Deployments. (Vgl. Axelos 2019, S. 161) Es nimmt explizit Bezug auf Continuous Delivery, also das fortlaufende Integrieren, Testen und Ausliefern von einzelnen Komponenten, das im Grunde nur bei Verwendung agiler Methoden in der Softwareentwicklung sinnvoll ist. Für seinen Erfolg bedarf es der Entscheidung für den Ansatz *Direct Integration*, der ein Produktivdeployment bei Fertigstellung erlaubt, aber umfangreiche Testautomatisierung voraussetzt (vgl. Axelos 2020, S. 31) Nur der Vollständigkeit halber sei erwähnt, dass ITIL schon immer die Beteiligung der Entwickler am Deployment in Produktivumgebungen unterstützt hat. ITIL4 erwähnt nun auch ausdrücklich die Möglichkeit, dass es sich dabei um Externe oder um Serviceintegratoren handelt. (vgl. Axelos 2019, S 161)

Die ITIL4 Foundation führt sieben universelle Leitprinzipien ein, die eine ITIL-basierte IT-Organisation bestimmen:

- Konzentration auf Wertschöpfung,
- Start mit dem Ist-Zustand,
- iterativer Fortschritt mit Feedback,
- transparente Zusammenarbeit,
- Ganzheitlichkeit,
- Einfachheit sowie
- Optimierung und Automatisierung. (vgl. Axelos 2019, S 39)

Ein Jahr später wird ausführlich beschrieben, wie diese Prinzipien mit Continuous Delivery in Bezug gebracht werden können:

- Wertschöpfende Entwicklung früher zum Kunden bringen.
- Inkremente werden zum bereits bestehenden Ist-Zustand hinzugefügt.
- Kurze Feedback-Schleifen sind Kern agilen Arbeitens.
- Transparenz der Arbeitsergebnisse für alle Beteiligten, also Entwicklung, Betrieb und Kunden.
- Die Ziele des IT-Dienstleisters und des Kunden werden holistisch betrachtet.
- Auf unproduktive Aktivitäten wird nach Möglichkeit verzichtet.
- Auslieferung von Code und Feedback sollen automatisiert werden. (vgl. Axelos 2020, S.45)

Die offensichtliche Nähe der neuen Leitprinzipien zu Agilität und Continuous Delivery lässt vermuten, dass entgegen der Reihenfolge der Veröffentlichungen diese originär Leitbild bei der Entstehung von ITIL4 waren. Gleichwohl wird betont, dass diese Lösung nicht in jedem Fall anwendbar ist. Sie passt insbesondere dann nicht, wenn gegenwärtige und zukünftige Anforderungen besonders unklar sind. (vgl. Axelos 2020, S. 46)

Die Schnittstelle zwischen Entwicklungsprojekten einerseits sowie Systembetrieb und Benutzersupport andererseits birgt seit jeher großes Konflikt- und Fehlerpotential. DevOps ist angetreten, dieses abzumildern und schlägt dazu u.a. vor, die strenge organisatorische Trennung beider Zuständigkeiten aufzulösen. Schon in ITIL V3 finden sich Ansätze dazu insbesondere im Prozess *Application Management*: „Umfassendes Application Management hebt das Nebeneinander von Software-Entwicklung und Service Management auf, indem es die Software-Entwicklungsphasen und die ... Service Management-Phasen in einen einzigen Lebenszyklus integriert.“ (Ebel, S. 324)

Wie bereits ausgeführt, kann die ITIL4-Practice *Monitoring and Event Management* eine der wesentlichen Voraussetzungen für erfolgreiche DevOps-Organisationen bereitstellen: eine Telemetrie-Infrastruktur. Zunächst sind die zu monitorierenden Services und Infrastrukturkomponenten zu identifizieren, dann sowohl deren native als auch die werkzeuggestützten Monitoringfunktionen einzurichten, Metriken und Schwellenwerte festzulegen, vorab die Reaktion auf bestimmte Events zu bestimmen und die zugehörigen Prozesse ein-

zusetzen. (vgl. Axelos 2019, S. 118f.) Ausdrücklich erwähnt wird der Einbezug von Betriebs- und Applikationsverantwortlichen: „For the definition of monitoring strategies and specific thresholds and assessment criteria, it can help to bring in a broad range of perspectives, including infrastructure, applications,...“ (Ebenda, S. 129) Technisch wird diese holistische Sicht auf Erfassung und Auswertung von Monitoring-Informationen ermöglicht, indem anfallende Daten zentral erfasst werden. Auf diesen Daten wird dann algorithmisch nach Mustern und Auffälligkeiten gesucht. (vgl. Axelos 2020, S. 40)

### Fazit

Es wurde deutlich, dass alleine die Existenz funktionierender ITSM-Prozesse nach ITIL der Arbeitsweise nach DevOps nicht im Wege steht. Wenn dieser Eindruck entsteht, dürfte er auf eine Überformalisierung zurückzuführen sein, in der insbesondere die Prozesse für Änderungskontrolle und Konfigurationsmanagement zu bürokratisch mit Leben gefüllt werden. Beide bieten die Flexibilität, DevOps-kompatibel gestaltet zu werden. Ein Konfigurationsmanagement, das gewissenhaft die DML aktuell hält, ist umgekehrt sogar Voraussetzung für agiles Deployment, ebenso ein funktionierendes Event Management, das Monitoringdaten verarbeitet.

Sowohl ITIL als auch DevOps fordern gemeinsame Ziele für Entwicklung und Systembetrieb sowie die dazu passenden Qualifikationen der Beteiligten.

Diese Synopse der beiden Konzepte liefert als Erkenntnis, dass ITIL-Prozesse ohne DevOps-Prinzipien möglich sind, dass umgekehrt aber DevOps ohne ITIL oder vergleichbare ITSM-Prozesse kaum erfolgreich in größeren Organisationen sein wird. Wesentliche Bedingung für die Kompatibilität von ITIL und DevOps ist ein flexibler und unbürokratischer Umgang mit ITSM-Praktiken, insbesondere mit Change Control. Wünschenswert bleibt für die Zukunft eine empirische Untermauerung dieses Ergebnisses anhand von Beispielen aus der Praxis.

### LITERATUR

- Alt, Rainer; Auth, Gunnar; Kögler, Christoph. 2017. Innovationsorientiertes IT-Management mit DevOps, Wiesbaden: Springer
- Andenmatten, Martin. 2014: Wie Cloud Services das Betriebsmodell in Unternehmen verändert“. In: HMD Praxis der Wirtschaftsinformatik 51.6 (2014), S. 782–794
- Axelos (Hrsg.) 2019. ITIL Foundation. ITIL4 Edition, Norwich: The Stationery Office, 2019
- Axelos (Hrsg.) 2020. ITIL4: Create, Deliver and Support, Norwich: The Stationery Office, 2020
- Betz, Charles. 2017. ITIL and DevOps: Differences & Frameworks Working Together – BMC Blogs. 2017. <https://www.bmc.com/blogs/itil-and-devops-lets-not-paper-over-the-differences>
- Böttcher, Roland. 2008. IT-Service Management mit ITIL V3, Hannover 2008
- Corless, Barry. 2017. ITIL® Retains Rank Among Top 5 Most Popular and Top 20 Highest-Paying IT Certi-

fications, 2017.  
<https://www.globalknowledge.com/us-en/resources/resource-library/articles/itil-retains-rank-among-top-5-most-popular-and-top-20-highest-paying-it-certifications/>

Ebel, Nadin. 2008. Itil V3 Basis-Zertifizierung, München 2008

Kim, Gene; Debois, Patrick; Willis, John; Humble, Jez. 2017. Das DevOps Handbuch: Teams, Tools, und Infrastrukturen erfolgreich umgestalten. 1. Auflage. Heidelberg: dpunkt.verlag

## **KONTAKT**

Prof. Dr. Carsten Dorrhauer  
Hochschule Ludwigshafen  
Ernst-Boehe-Straße 4  
D-67059 Ludwigshafen  
[carsten.dorrhauer@hwg-lu.de](mailto:carsten.dorrhauer@hwg-lu.de)  
+49 621 5203 330

# Die Smart Factory und ihre Risiken

Frank Herrmann

Ostbayerische Technische Hochschule Regensburg (OTH Regensburg)  
Innovationszentrum für Produktionslogistik und Fabrikplanung  
Postfach 120327  
93025 Regensburg  
E-Mail: Frank.Herrmann@OTH-Regensburg.de

## Schlüsselwörter

Industrie 4.0, Smart Factory, Risiken

## Abstrakt

In dieser Arbeit werden die Risiken einer Smart Factory untersucht und dargestellt, um den Status der Realisierung von Smart Factories aufzuzeigen. Dazu werden im Detail die entscheidenden technischen Komponenten vorgestellt und im Hinblick auf ihre Risiken analysiert. Generell wird eine ganzheitliche Sichtweise auf eine Smart Factory eingenommen. Ganzheitlich werden die ökonomischen Risiken erhoben. Diese komplette Risikoanalyse zeigt, dass der größte Handlungsbedarf im technologischen Bereich liegt. Dabei wurden die Themen Standardisierung, Informationssicherheit, Verfügbarkeit der IT-Infrastruktur, Verfügbarkeit von schnellem Internet und komplexen Systemen als am Wichtigsten eingestuft, und diese werden im Detail weiter analysiert. Ebenso weiter analysiert werden die organisatorischen und finanziellen Risiken, welche ebenfalls eine wichtige Rolle bei einer Transformation in eine Smart Factory spielen.

## 1. Einleitung

Um den Kundenerwartungen in der heutigen Zeit gerecht zu werden, bedarf es selbst bei Losgröße eins einer flexibleren Produktion. Um dieses Ziel zu erreichen, sollte eine neue Stufe der Automatisierung durch die Einführung von Methoden der Selbstoptimierung, Selbstkonfiguration, Selbstdiagnose, Kognition und intelligenten Support der Arbeitnehmer bei ihrer zunehmend komplexen Arbeit realisiert werden. Dies wird unter dem Begriff "Smart Factory" oder „Industrie 4.0“ zusammengefasst. In einer traditionellen Fabrik ist die Bereitstellung von qualitativ hochwertigen Dienstleistungen oder Produkten mit möglichst geringen Kosten der Schlüssel zum Erfolg. Industriefabriken versuchen hierbei, so viel Leistung wie möglich zu erzielen, um sowohl ihren Gewinn zu steigern als auch ihren Ruf zu verbessern. Im Gegensatz dazu können Komponenten und Systeme in einer Fabrik der Industrie 4.0 zusätzlich zur Zustandsveränderung, Fehlerdiagnose, Selbstwahrnehmung und Selbstvorhersage lernen, wodurch dem Management ein besserer Einblick in den Status der Fabrik ermöglicht wird.

Darüber hinaus ermöglicht der Peer-to-Peer-Vergleich und die Verbindung von Zustandsinformationen aus verschiedenen Komponenten eine präzise Zustandsprognose auf Komponenten- und Systemebene und zwingen das Fabrikmanagement, die erforderliche Wartung zum bestmöglichen Zeitpunkt auszulösen, um eine Just-in-Time-Wartung zu ermöglichen und Ausfallzeiten nahe Null zu erreichen. Es existieren bereits viele Publikationen über die intelligente Fabrik – Literaturüberblicke sowie Analysen von Forschungsarbeiten. Diese Publikation erweitert die Literaturüberblicke, indem die Risiken einer Smart Factory, die aus eben den vorhandenen Publikationen gefolgert wurden, zusammengefasst und bewertet werden. Möglicherweise steigt dadurch das Bewusstsein für diesen Fertigungstrend – vor allem für Anwender.

Der Rest dieser Arbeit ist wie folgt strukturiert. Zunächst werden Smart Factory bzw. Industrie 4.0 vorgestellt. Als nächstes werden die Technologien innerhalb einer Smart Factory erläutert. Dabei wird besonders auf Künstliche Intelligenz eingegangen. In Abschnitt 4 werden die Risiken, basierend auf einer Literaturübersicht, für diese Technologien aufgelistet und analysiert. Abschließend werden einige Schlussfolgerungen gezogen.

## 2. Industrie 4.0

In der industriellen Revolution, s. auch die Abbildung 1, brachte die Dampfmaschine, das Fließband und der Computer tiefgreifende wirtschaftliche Veränderungen. Nun stehen wir vor dem nächsten gewaltigen industriellen Wandel, der sogenannten 4. Industriellen Revolution oder kurz, Industrie 4.0.

Die Idee der vierten industriellen Revolution besteht darin, dass sich die Produktion mit modernsten Informations- und Kommunikationstechniken verzahnt. So entsteht die Möglichkeit, Produkte nach individuellen Kundenwünschen zu fertigen und diese in Losgröße eins zum Preis von Massenware zu produzieren. Die technischen Grundlagen bilden intelligente, digital vernetzte Systeme und Produktionsprozesse. Des Weiteren bestimmt die Industrie 4.0 die gesamte Lebensphase eines Produktes. Sie befasst sich mit der Idee, der Entwicklung, der Fertigung, der Nutzung und der Wartung bis hin zum Recycling des Produktes (Bundesministerium für Wirtschaft und Energie).

Alle vier Revolutionsstufen haben eine Gemeinsamkeit. Sie haben alle eine Auswirkung auf das wirtschaftliche und das gesellschaftliche Leben. Ebenfalls steigt die Produktivität an und schafft Veränderung in den Arbeitsprozessen und Arbeitsbedingungen. Für die Industriemitarbeiter bedeutet dies, eine höhere Qualifikationsanforderung. Mit diesen Veränderungen muss auch in der bereits anstehenden vierten Revolution gerechnet werden; s. (Kelkar et al. 2014, S. 2).

Unternehmen einen hohen Digitalisierungsgrad ihrer Wertschöpfungskette. Aktuell nutzen bereits 20 Prozent der deutschen Unternehmen selbststeuernde Anlagen; s. (Bundesministerium für Wirtschaft und Energie).

Durch die stärkere, innerbetriebliche sowie überbetriebliche Vernetzung, wird eine völlig neue Art der wirtschaftlichen Produktion geschaffen. Dies stellt deutsche und weltweite Unternehmen vor großen Herausforderungen, da Industrie 4.0 alle Dimensionen, wie Technik, Or-

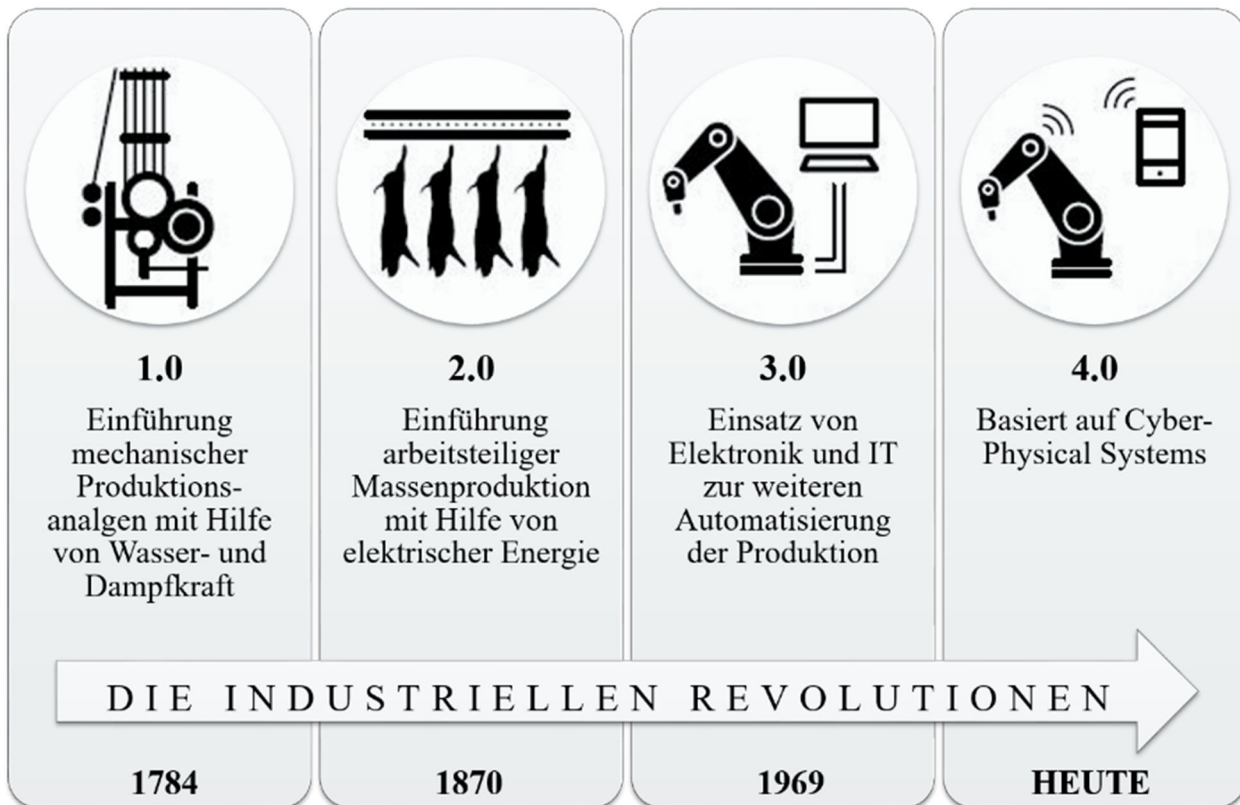


Abbildung 1: Industrielle Revolutionen im Überblick. (Eigene Darstellung)

Den Herausforderungen der Industrie 4.0 und der Umstellung der Produktion auf einen vernetzten Betrieb müssen sich deutsche Unternehmen stellen. Gelingt dies nicht, so besteht die Gefahr, dass deutsche Unternehmen den wirtschaftlichen Anschluss verlieren; s. (Bischoff 2015, S. 3).

Werden jedoch aktuelle Zahlen der Industrie 4.0 für den Standort Deutschland betrachtet, so lässt sich eine positive Entwicklung hin zur Industrie 4.0 erkennen. Laut dem Bundesministerium für Wirtschaft und Energie werden bis zum Jahr 2020 jährlich Investitionen von 40 Milliarden Euro in Industrie 4.0 Anwendungen geplant. Ebenfalls wird ein zusätzliches volkswirtschaftliches Wachstum von über 153 Milliarden Euro bis 2020 erwartet. Bis zum Jahr 2020 sehen 83 Prozent der deutschen

Unternehmen, Menschen und Geschäftsmodelle, betrifft; s. (Bischoff 2015, S. 1).

Um die Risiken und Herausforderungen für Unternehmen in der Industrie 4.0 genauer betrachten zu können, wird nun vorab die moderne Produktionsumgebung mit dem aktuellen technischen Stand beschrieben.

### 3. Technologien innerhalb einer Smart Factory

Im Zentrum der Industrie 4.0 steht die Smart Factory (intelligente Fabrik). Grundsätzlich geht es bei der Smart Factory um die Vernetzung von Maschinen und Anlagen mittels Software, so dass eine intelligente Kommunikation miteinander ermöglicht wird und die Arbeitsschritte automatisiert aufeinander abgestimmt werden können. Damit dies realisiert werden kann, müssen grundlegende



Technologien eingesetzt werden. Die technologischen Voraussetzungen werden in diesem Kapitel erläutert; s. (Zillmann 2016, S. 8).

Um die technischen Voraussetzungen einer Smart Factory aufzuzeigen, müssen vorerst die Cyber-Physical Systems (CPS), welche die Grundlage einer Smart Factory bilden, beschrieben werden; s. (Bauernhansl et al. 2014, S. 15f.). Das Schlagwort CPS tauchte erstmals in Amerika im Jahr 2006 auf. Systeme erhalten den Zusatz „Cyber“, wenn sie zur diskreten Verarbeitung von Informationen und zur Kommunikation genutzt werden. Daneben werden die realen Systeme als „physical“ bezeichnet. Dies bedeutet, die CPS sind autonom und können sich eigenständig konfigurieren und erweitern; s. (Jasperneite 2012). Ein weiteres wesentliches Merkmal der CPS ist, dass sie sich über offene und globale Informationsnetze, wie das Internet, verbinden können. Somit können sich Systeme beliebig koppeln, ihre Verbindungen während der Betriebszeit verändern, beenden und neu aufbauen. Es entsteht die Möglichkeit, verfügbare Daten, Informationen und Dienste an beliebiger Stelle im CPS bereitzustellen und zu verwenden; s. (VDI/VDE-Gesellschaft 2013, S. 2). CPS können Objekte, Produkte, Geräte, Gebäude, Produktionsanlagen oder Logistikkomponenten sein, die eingebettete Systeme enthalten; s. (Bauernhansl et al. 2014, S. 15f.).

*Eingebettete Systeme* bzw. Embedded Systems sind Mikroprozessoren oder kleine Rechner. Sie sind ein zentrales Thema bei der Umsetzung von CPS. Durch die Kombination von Hardware- und Softwarekomponenten wird das entsprechende System gesteuert, geregelt oder überwacht. Mit Hilfe dieser Technologie entstehen die oben beschriebenen flexiblen und intelligenten Systeme (CPS); s. (Mauerer 2016).

Die Embedded Systems sind mit *Sensoren* und *Aktoren* ausgestattet. Über Sensoren werden die Produktionsdaten erfasst. Die gesammelten Daten werden mit verfügbaren Daten und Diensten ausgewertet und abgespeichert. Durch die Verwendung von Aktoren besitzt ein CPS die Möglichkeit, physikalisch auf seine Umwelt einzuwirken; s. (Bauernhansl et al. 2014, S. 15f.). Die Kommunikation der CPS erfolgt über das Internet oder Intranet. Sie können Internetdienste nutzen und / oder diese selbst anbieten.

Die Bezeichnung *Internet of Things* (IoT) wurde im Jahr 1999 zum ersten Mal verwendet. Das klassische Internet wurde umfassend erweitert. Dies hat zur Folge, dass die virtuelle und reale Welt durch die Vernetzung von und mit Alltagsgegenständen, immer mehr miteinander verschmelzen. Physische Objekte (things) werden eindeutig identifizierbar, erhalten eine virtuelle Repräsentation und können über das Internet kommunizieren; s. (Mauerer 2016). Die Kommunikation erfolgt über offene und globale Informationsplattformen. Über diese Netze werden Dienste und Daten genutzt und zur Verfügung gestellt. Solche Dienste werden in der Literatur als *Internet of*

*Things and Services* (IoTS) bezeichnet. Es entsteht die Vision von adaptiven und flexiblen Produktionsanlagen, die sich selbst konfigurieren und teilweise selbst organisieren; s. (VDI/VDE-Gesellschaft 2013, S. 3).

Zur Steuerung von intelligenten Produktionsanlagen sind innerhalb einer Produktionsumgebung *Cyber Physical Production Systems* (CPPS) notwendig. Die Aufgabe des CPPS besteht darin, die einzelnen CPS zu koordinieren und die Veränderungen in der Produktion zu steuern (Huber 2016a, S. 39). Mit Hilfe dieser Plattform können die drei verschiedenen Internetarten miteinander verbunden werden. Im Internet des Menschen vernetzt sich der Mensch mit den sozialen Netzwerken. Des Weiteren gibt es das zuvor beschriebene Internet der Dinge (IoT) und das Internet der Dinge und Dienste (IoTS); s. (Bauernhansl et al. 2014, S. 16).

Auch wenn das Ziel der Smart Factory eine vernetzte, selbstorganisierte Produktionsumgebung ist, darf der Mensch in diesem Szenario nicht fehlen. Der Mensch muss weiterhin im Mittelpunkt stehen, wobei sich sein Aufgabengebiet allerdings sehr stark ändern wird. Hauptsächlich müssen Menschen kontrollierende Tätigkeiten wahrnehmen, während die operativen Aufgaben in den Hintergrund treten. Für die Ausführung dieser Tätigkeiten muss der Mensch über eine multimodale Schnittstelle, auch *Human-Machine-Interface* (HMI) genannt, mit den CPS verbunden sein; s. (Huber 2016a, S. 286). Ein CPS kann beispielsweise über Sprache oder Touch Displays gesteuert werden. Zukünftig kann der Mensch auch über Gesten entsprechend einwirken; s. (Bauernhansl et al. 2014, S. 16). Durch mobile Endgeräte wird der Mensch direkt in das Kommunikationsnetz der Smart Factory eingebunden; s. (Zillmann 2016, S. 13). Geeignete mobile Endgeräte sind Tablets oder Smartphones, wodurch die intuitive Benutzung dieser Devices eine Vereinfachung für den Mitarbeiter darstellt. Das Schlagwort *Augmented Reality* (AR) dient als Hilfsmittel für den Produktionsmitarbeiter. Augmented Reality bildet die reale Welt mit zusätzlichen Informationen aus der virtuellen Welt ab. Die angezeigten Informationen werden in der richtigen Größe und an der korrekten Position dargestellt. Diese neue Technologie wird in Smart Glasses oder in Datenhandschuhen verwendet, welche mit dem Backend-System verbunden werden; s. (Huber 2016a, S. 78).

Beim Thema CPS muss auch die *Machine-to-Machine-Communication* (M2M-Kommunikation) erwähnt werden. Hier kann nicht nur der Mensch mit der Maschine kommunizieren, sondern die Maschine kann selbstständig mit anderen Maschinen oder Werkzeugen in Kontakt treten. Durch diese Lösung können Maschinen- und Produktionsdaten erfasst, und in Echtzeit weitergeleitet werden; s. (Zillmann 2016, S. 13).

Die Echtzeitverarbeitung von Informationen wird auch *Real Time Enterprise* (RTE) genannt und ist ein zentrales

Element innerhalb einer Smart Factory. Die Informationen sollten dem Anwender immer rechtzeitig vorliegen. Bei der Verarbeitung großer Datenmengen spielt das Schlagwort Big Data eine wichtige Rolle; s. (Huber 2016a, S. 11).

*Big Data* wird oftmals als „Rohstoff des 21. Jahrhunderts“ bezeichnet. Die Grundvoraussetzung hierbei ist das Sammeln, Speichern, Filtern, Analysieren, Verdichten und Visualisieren großer Datenmengen. Es kann sowohl mit strukturierten, als auch mit unstrukturierten Daten gearbeitet werden. Die Hauptaufgabe ist es, Muster und Kausalitäten zu erkennen. Bei dem richtigen Einsatz von Big Data zeigt sich, dass mit zunehmender Datenmenge auch die Aussagekraft steigt.

Technisch können Big Data-Systeme auf Not Only SQL (NoSQL)-Datenbanken oder auf einem Hadoop Framework basieren. Hadoop ist ein Open-Source-Framework und dient zur Speicherung großer Datenmengen auf verteilten Systemen. Ein weiteres Big Data-System ist die In-Memory-Datenbank. Diese wird für schnelle Auswertungen eingesetzt. Hier dient der Arbeitsspeicher eines Computers als Datenspeicher. Die SAP SE ist einer der bekanntesten In-Memory-Datenbank-Anbieter; s. (Huber 2016a, S. 54ff.).

Damit eine gemeinsame Vernetzung und Kommunikation zwischen verschiedenen Umgebungen und Objektarten erfolgen kann, müssen akzeptierte *Standards und Schnittstellen* eingesetzt werden. Dazu wurde im Jahr 2012 das Internet-Protokoll IPv6 eingeführt. Mit Hilfe des IPv6 Protokolls stehen für Ressourcen, Informationen, Objekte und Menschen in einer Smart Factory eine ausreichende Anzahl an Adressen zur Verfügung. Es entsteht somit eine flächendeckende Vernetzung; s. (Kelkar et al. 2014, S. 8).

Innerhalb einer Smart Factory bilden *Cloud-Computing-Lösungen* eine wichtige Komponente. Hierbei wird eine bedarfsgerechte und flexible Nutzung von unterschiedlichen IT-Leistungen über das Internet in Form von Diensten angeboten. Allgemein betrachtet werden bei einem Cloud-Ansatz Daten oder Dienste an einen internen oder externen Dienstleister ausgelagert; s. (Huber 2016a, S. 60g.). Die Abrechnung solcher Services erfolgt immer nutzungsabhängig. Es können drei Service-Elemente unterschieden werden. Unter *Infrastruktur-as-a-Service (IaaS)* versteht man die Bereitstellung von Rechen-, Speicher- und Netzkapazität durch einen Cloud-Provider. Bei dem Service-Element *Platform-as-a-Service (PaaS)* wird dem Anwender eine Entwicklungsumgebung in Form eines technischen Frameworks über einen Internetanbieter zur Verfügung gestellt. Hierbei handelt es sich um standardisierte Umgebungen. Diese vereinfachen und beschleunigen die Softwareentwicklung, indem die angebotene Umgebung nicht implementiert werden muss. *Software-as-a-Service (SaaS)* stellt standardisierte Anwendungsservices über eine Cloud zur Verfügung und

kann von Endanwendern genutzt werden; s. (Münzl et al. 2015, S. 10ff.).

Durch die vorangegangenen Definitionen wird deutlich, dass ohne *Automatisierung* eine Produktion im Sinne einer Smart Factory nicht erfolgen kann. Mittels der Automatisierung werden Produktionsprozesse beschleunigt und optimiert. Dies geschieht durch die Übertragung der Produktionsprozessfunktionen an künstliche Systeme. Als Automat wird eine Maschine bezeichnet, welche bestimmte und vorher festgelegte Abläufe selbstständig beziehungsweise automatisch ausführt. Je nach Grad der Automatisierung wird von Teil- oder Vollautomatisierung gesprochen. Im Bereich der Smart Factory wird eine Vollautomatisierung gefordert. Dies bedeutet, dass keine vollständige Weitergabe von starren und wiederkehrenden Produktionsabläufen an Maschinen erfolgt. In einer Smart Factory werden die Prozesse mit unterschiedlichen Aufgabenstellungen durch flexible Fertigungssysteme ausgeführt; s. (Maurer 2016).

Die *sensitiven Roboter* gelten als Inbegriff der Automatisierung und sind eng mit den Themen CPS und M2M-Kommunikation verknüpft. Ein sensitiver Roboter wird mit feinfühligem Gelenksensoren ausgestattet, um somit Tätigkeiten in Verbindung mit menschlichen Kollegen durchzuführen. Innerhalb einer Smart Factory bewegen sich die Roboter frei und werden nicht durch Gitterzäune vom Menschen getrennt. Diese Zusammenarbeit wird *Mensch-Roboter-Kollaboration* bezeichnet; s. (Reinhart, Gunther et al. 2018). Der Mensch wird durch die enge Zusammenarbeit von körperlich schweren Arbeiten entlastet oder bei ergonomisch ungünstigen Tätigkeiten unterstützt; s. (Huber 2016a, S. 50).

Als *digitales Produktgedächtnis* wird ein zu fertigendes Produkt mit einem automatisch auslesbaren Datenträger bezeichnet, welches dazu beiträgt, ein vollständig dezentralisiertes Produktionsumfeld zu erreichen. Somit werden die Produkte selbst zum Informationsträger. Die relevanten Daten werden unabhängig von Server-Systemen an die bearbeitenden Maschinen weitergegeben. Das digitale Produktgedächtnis ist für die autonome, dezentrale Produktion und für individualisierten Produkte ein entscheidender Schritt. Es ermöglicht eine durchgängige Produktdokumentation über die gesamte Produktlebensphase. Die Produktdaten sind jederzeit auslesbar und bei Problemen können fehlerhafte Bauteile sofort identifiziert werden. Durch das digitale Produktgedächtnis wird eine optimale Ressourcenoptimierung ermöglicht; s. (Kelkar et al. 2014, S. 8).

Im Zusammenhang einer flexiblen und effizienten Produktion muss die neue Technologie des *3D-Druckers* erwähnt werden. Durch diese Technik werden additive Fertigungsverfahren ermöglicht. Additiv bedeutet in diesem Umfeld die Erzeugung eines Bauteils über schichtweises Hinzufügen eines Werkstoffes. Das gewünschte Objekt wird Schicht für Schicht auf Basis eines 3D-Modells aus

den gewünschten Materialien erzeugt. Die Vorteile gegenüber einer konventionellen Fertigung sind enorm. Kleinserien können wirtschaftlich und rentabel produziert werden und die Herstellung schwieriger Bauteile wird erleichtert. Ein weiterer positiver Aspekt ist die Minimierung des Materialeinsatzes und das Entfallen der Fertigungsschritte. Das Produktionsumfeld der Smart Factory lässt sich mit Hilfe dieser neuen Technologie einfacher und flexibler umsetzen und gestalten; s. (Huber 2016a, S 33ff.).

In einer Smart Factory haben *robuste Netze* eine hohe Bedeutung. Damit die CPS miteinander kommunizieren können, ist eine hohe Verfügbarkeit der kabel- und funkgestützten Netzwerke unabdingbar. Die Herausforderung der Datenfülle und der damit verbundenen hohen Datenübertragungsrate darf nicht unterschätzt werden. Um dieses Problem zu bewältigen, werden Breitbandnetzwerke eingesetzt. Eine besondere Rolle spielen Wireless LANs (Local Area Network), so dass die Mobilität von Robotern und Automaten gewährleistet werden kann. Im Bereich Funknetzwerken besteht derzeit noch Entwicklungsbedarf (Kelkar et al. 2014, S. 9).

Durch diese Technologien hat eine Smart Factory ein komplexes Umfeld und es treten hohe technischen Anforderungen auf. Diese Überlegungen werden durch die Abbildung 2 vervollständigt. Sie dient zusätzlich der Einordnung der Begrifflichkeiten, die üblicherweise mit der Smart Factory in Verbindung gebracht werden. Einige

der genannten Technologien existieren bereits und manche befinden sich noch in der Entwicklung. Dennoch sind die genannten Technologien und Methoden noch nicht in dem gewünschten Ausmaß und Zusammenspiel eingesetzt. Eine zentrale Rolle für die Umsetzung der Zusammenarbeit der einzelnen Technologien spielt die IT. Sie sorgt für die notwendige Vernetzung und die benötigten Architekturen. Erst durch den richtigen Einsatz der IT werden all diese Komponenten steuerbar. Durch die ansteigende Vernetzung der Technologien treten viele Risiken auf und erfordern neue Konzepte und Mechanismen in Unternehmen. Welche Risiken auftreten können, werden im Kapitel „Risiken innerhalb einer Smart Factory“ erläutert. Zuvor wird auf immer intensiver genutzte Künstliche Intelligenz in einer Smart Factory näher eingegangen.

#### 4. Künstliche Intelligenz

Industrie 4.0 ist die erste industrielle Revolution, die nicht nach ihrem Ende definiert wurde, sondern während ihrer Laufzeit antizipiert wird. Sie lässt sich in drei aufeinander aufbauenden Stufen einteilen, welche im Folgenden beschrieben sind. Der Bereich Künstliche Intelligenz bildet mit drei Phasen die letzte Form der industriellen Revolution, s. (Botthof, Alfons et al. 2015, S. 24f.): Die erste Stufe innerhalb der Industrie 4.0 ist die Flexibilisierung. Cyber-physische-Systeme vernetzen hierbei die Prozesse der Produktion, Logistik, Internetdienste, des Engineerings und des Managements miteinander.

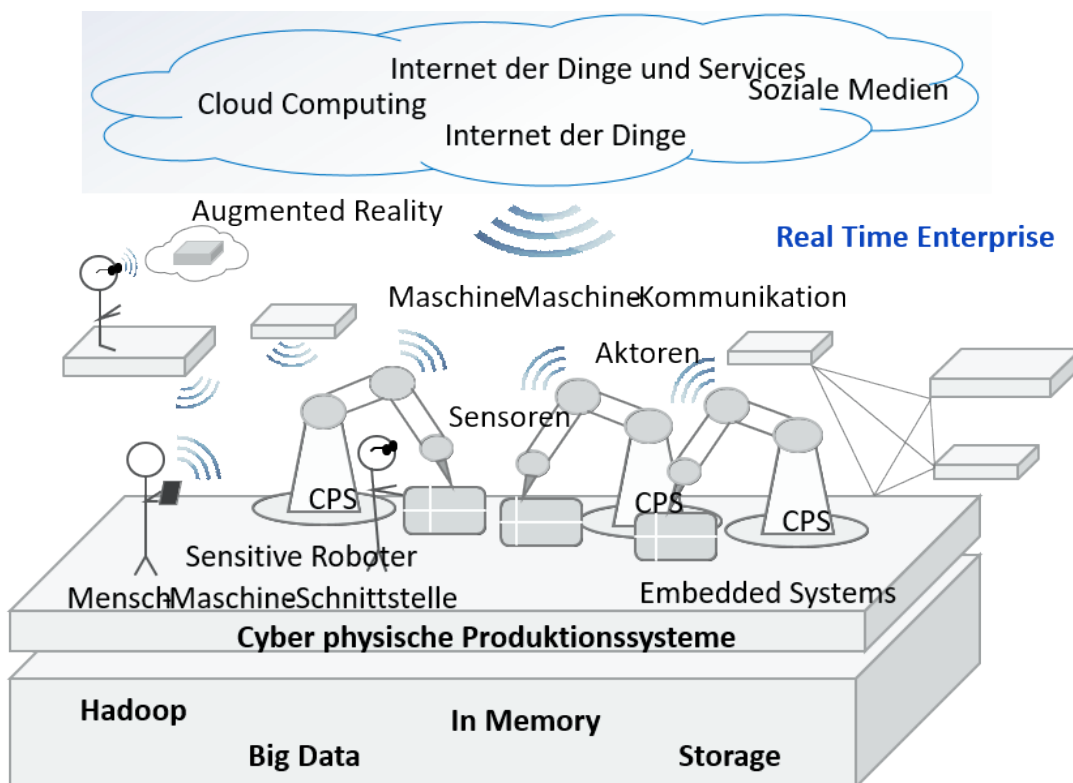


Abbildung 2: Begrifflichkeiten innerhalb einer Smart Factory. (Eigene Darstellung)

Dadurch werden Informationen selbstständig mittels Sensoren erfasst und ausgetauscht. Die transferierten Informationen ermöglichen den Start von Aktionen und somit eine eigenständige und gegenseitige Steuerung. Allerdings handelt es sich an dieser Stelle noch nicht um eine Digitalisierung, jedoch ist Vernetzung die Grundlage dafür; s. (Ruskowski, 2019).

Die Humanisierung bzw. Kognition stellt die zweite Ebene dar. Die bestehenden cyber-physischen Systeme werden humanisiert, indem sie mit kognitiven Eigenschaften wie Sehen und Hören erweitert werden. Dies kann durch Bilderkennungssysteme in der Produktion erfolgen, die beispielsweise dabei helfen, Ware mit Qualitätsmängeln zu identifizieren. Ein weiteres Beispiel der Kognition ist der Einsatz von Spracherkennungssystemen, die eine Steuerung der Maschinen mittels der Stimme ermöglichen; s. (Ruskowski, 2019).

Der letzten Stufe der Industrie 4.0, der Agilität, wird die Künstliche Intelligenz zugeordnet. Bei der autonomen Herstellung der Produkte durchlaufen Bauteile die einzelnen Produktionsstufen ohne menschliche Hilfe. Durch den Schritt der Humanisierung der Maschinen wissen diese, um welches Bauteil es sich handelt und welche Arbeitsschritte an ihm ausgeführt werden sollen. Das führt schließlich dazu, dass Produktion und Logistik durch die ganzheitliche Einführung von KI ineinander einfließen; s. (Ruskowski, 2019).

Dennoch: eine allgemeingültige Definition des Begriffs „Künstliche Intelligenz“ existiert auf Grund seiner Anpassung an die technischen Möglichkeiten seit der Begriffsbildung in den 1950er Jahren nicht. Das Fraunhofer Institut beschreibt Künstliche Intelligenz in einer Studie als „ein Teilgebiet der Informatik, das versucht, mit Hilfe von Algorithmen kognitive Fähigkeiten wie Lernen, Planen oder Problemlösen in Computersystemen zu realisieren.“; s. (Bauer et al. 2018, S. 11) Somit handelt es sich hierbei um Technologien, welche gewisse Ausprägungen der menschlichen Intelligenz adoptieren (Krüger et al., 2019). Die Merkmale dieser Systeme lassen sich in die vier folgenden Kernbereiche unterteilen; s. (Termer, 2019):

**Wahrnehmen** – In Unternehmen müssen IT-Systeme die Daten der verschiedenen Abteilungen aufnehmen, um diese in den weiteren Schritten verarbeiten zu können. Durch den weiter steigenden Fortschritt der Digitalisierung sind der Menge und der Art von Daten kaum Grenzen gesetzt. Einerseits stützt dies die Zuverlässigkeit der Aussage, welche das Ergebnis der Datenverarbeitung liefert, andererseits haben nicht alle Daten die gleiche Relevanz für die entsprechenden Systeme. Daher ist es eine gründliche Filterung der Daten notwendig, welche für den Verarbeitungsprozess erforderlich sind.

**Verstehen** – Die wahrgenommenen Daten sollen automatisiert erkannt und zugeordnet werden. Dies kann mittels

Erweiterungen durch Deep Learning bzw. Machine Learning realisiert werden. Immer weiter steigt hierbei die Nachfrage an Sprachverständnis-Tools, die beispielsweise bei der Bearbeitung von Kundenanliegen als Unterstützung dienen. Man betrachte nun ein hypothetisches Unternehmen, welches täglich eine hohe Zahl an Produktionsaufträgen erhält und diese automatisiert an die betreffenden Produktionsabteilungen weiterleiten möchte. Wird eine definierte Abteilung in dem Auftrag erkannt, so kann dieser problemlos zugeordnet werden. Wenn allerdings ein neuer Auftrag keine bekannte Struktur aufweist, da beispielsweise eine Sonderbestellung hinzugefügt wurde, muss die KI zunächst den Wunsch verstehen können. Trainierte Systeme erfüllen bereits heutzutage diese Aufgabe.

**Handeln** – KI-Systeme ermöglichen eine intelligente Prozess- bzw. Gerätesteuerung in der Produktion. Auswirkungen eben solcher intelligenter Prozess- und Gerätesteuerungen auf den Herstellungsprozess, sollen unter anderem Diskussionsthema dieser Arbeit sein.

**Lernen** – In der Entwicklungsphase der Künstlichen Intelligenz werden dem System Daten injiziert, um bestimmte Strukturen zu erlernen. Der Lernprozess beschränkt sich jedoch nicht nur auf die sogenannte Trainingsphase, sondern streckt sich bis in den laufenden Betrieb. Folglich handelt es sich bei Künstlicher Intelligenz um ein System, welches durch ständige Datenerfassung kontinuierlich lernt und sich dadurch optimal an die Anforderungen der Unternehmen anpasst.

## 5. Risiken innerhalb einer Smart Factory

Da die Industrie 4.0 zunehmend an Form gewinnt, kristallisieren sich neben den vielen Vorteilen auch immer mehr Risiken heraus. Diese näher zu beleuchten ist essentiell, um im Nachgang den Stand der Smart Factory zu bewerten. Die Risiken zum Thema Industrie 4.0 können in die Bereiche technische, organisatorische und wirtschaftliche Risiken gruppiert werden. Da die Industrie 4.0 ein stark technologisch getriebenes Thema ist, werden im Folgenden die technologischen Risiken genauer betrachtet. Die organisatorischen und wirtschaftlichen Risiken werden zur Vollständigkeit danach nur kurz behandelt.

In den Smart Factories werden sehr viele Technologien eingesetzt, weswegen sich auf diejenigen beschränkt wurde, die in der Literatur am intensivsten behandelt werden und dadurch als am Wichtigsten eingestuft werden können. Sie lauten: Standardisierung, IT-Sicherheit, Verfügbarkeit der IT-Infrastruktur, Verfügbarkeit von schnellem Internet und komplexe Systeme beschränkt, da die in der Literatur am intensivsten behandelt werden und deswegen als am Wichtigsten eingestuft

### 5.1 Standardisierung

Zum Thema Industrie 4.0 existieren keine internationalen Standards und Normen. Diese sind jedoch Voraussetzung

für eine hohe Anzahl von Vernetzungspartnern und der Erschließung des wirtschaftlichen Potenzials der Industrie 4.0. Größtenteils passen sich die zuliefernden Unternehmen den Standards der größeren Unternehmen an, was zur Folge hat, dass die Handlungsmöglichkeiten der kleineren Unternehmen gehemmt werden. Ebenso entstehen zu hohe Investitionsausgaben, wenn Unternehmen auf nicht zukunftsorientierte Technologien setzen; s. (Schröder 2016, S. 12). Um eine Smart Factory zu verwirklichen und eine Grundlage für die Standardisierung vorzugeben, wurde ein Referenzarchitekturmodell für die Industrie 4.0 (RAMI 4.0) entwickelt; s. (Zillmann 2016, S. 11). Dieses wurde erstmals auf der Hannover Messe 2015 vorgestellt und von den Industrieverbänden Bitkom, VDMA und ZVEI gemeinsam definiert; s. (Schewe 2016). Anhand des Modells soll ein gemeinsames Verständnis geschaffen werden, welche Standards, Use Cases und Normen für eine Smart Factory notwendig sind. Das Ziel des Referenzarchitekturmodells ist es, möglichst wenig Standards einzusetzen und ein anschauliches und einfaches Architekturmodell als Referenz zu schaffen; s. (Plattform Industrie 4.0 2015, S. 41f.).

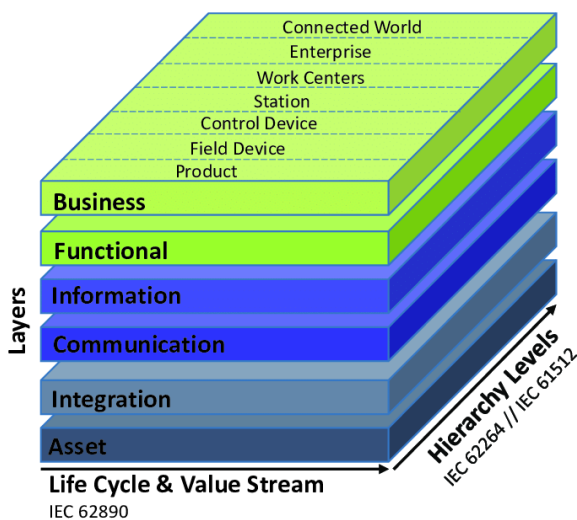


Abbildung 3: RAMI 4.0 (Eigene Darstellung)

Das Modell besteht aus einem dreidimensionalen Koordinatensystem. Die Achse „Hierarchy Levels“ stellen die unterschiedlichen Funktionalitäten und Verantwortlichkeiten innerhalb einer Fabrik oder einer Anlage dar. Die waagrechte Achse „Life Cycle & Value Stream“ bildet den Lebenszyklus und Wertschöpfungsketten von Anlagen und Produkten ab. Hier wird zwischen Typ und Instanz unterschieden. Ein Typ befindet sich in der Entwicklung oder Prototypenfertigung. Wenn das Produkt gefertigt wird, wird es als Instanz bezeichnet. Anschließend muss die senkrechte Achse „Layers“ betrachtet werden. Hier wird das digitale Abbild der betrachtenden Industrie 4.0-Technologie strukturiert Schicht für Schicht beschrieben. Es werden die sechs Schichten Geschäftsprozesse, Funktionen, Daten, Kommunikation, Integration und die realen Dinge dargestellt. Durch die drei beschriebenen Achsen können alle Aspekte von Industrie

4.0 abgebildet werden und dienen als Basis für weitere Arbeitsschritte; s. (Hankel und Rexroth 2015).

Die Hard- und Softwarekomponenten in der Produktion werden Industrie 4.0-fähig, wenn sie die Kommunikationsfähigkeit der realen Objekte mit den jeweiligen Daten und Funktionen umsetzen. Eine wichtige Voraussetzung dafür ist, dass die Industrie 4.0-Komponenten während ihres gesamten Lebenszyklus, die Daten in einem abgesicherten elektronischen Container sammeln und mit sich tragen. Der elektronische Container kann auch als Verwaltungsschale bezeichnet werden. Alle relevanten Daten einer Hard- oder Softwarekomponente sind in der Verwaltungsschale gespeichert, und ergeben zusammen ein virtuelles Abbild. Dadurch entstehen neue Möglichkeiten der vernetzten Produktion und liefern einen Mehrwert für alle, die am Wertschöpfungsprozess beteiligt sind.

Damit die Kommunikation entlang der Wertschöpfungskette funktioniert, ist die Open Platform Communications Unified Architecture, kurz OPC UA, eine grundlegende Voraussetzung. Dieses Kommunikationsprotokoll ist für M2M- und PC-to-Machine Kommunikation entwickelt worden und wurde im Jahr 2006, als Standardprotokoll im Rahmen der Industrie 4.0, vorgestellt. OPC UA kann Maschinendaten transportieren und diese ebenfalls maschinenlesbar semantisch beschreiben. Dies unterscheidet OPC UA von herkömmlichen Kommunikationsprotokollen; s. (Kagermann et al. 2013, S. 104).

Die aufgezeigten Beispiele zum Thema „Standardisierung“ sind nur ein kleiner Einblick und der Beginn für die Umsetzung einer vernetzten Smart Factory. Es besteht die dringende Notwendigkeit, internationale Standards und Normen in naher Zukunft zu erarbeiten, wobei auf die bereits bestehenden Technologien und Schnittstellen zurückzugreifen ist.

## 5.2 IT-Sicherheit

Eines der größten technischen Risiken innerhalb einer Smart Factory stellt die „IT-Sicherheit“ dar. In den letzten Jahren nahmen die Cyber-Attacken gegen Unternehmen sowie auch gegen Privatpersonen stetig zu; s. (Huber 2016b). Aufgrund des mangelnden Ausbaus von Datenschutz und Datensicherheit entstehen durch Cyber-Attacken hohe finanzielle Schäden. Auch die Ausbreitung neuer Geschäftsmodelle und der Einsatz von Cloud-Lösungen werden durch die aufkommenden Gefahren gehemmt; s. (Kelkar et al. 2014, S. 56). Mittels der CPS-basierten Produktionssysteme entstehen neue Sicherheitsanforderungen, da sie einen hohen Vernetzungsgrad aufweisen. Sie sind vernetzt mit einer Vielzahl an beteiligten Maschinen, IT-Systemen und Automatisierungskomponenten. Die technischen Systemkomponenten agieren teilweise autonom mit einem zeitkritischen Daten- und Informationsaustausch, wobei zugleich mehrere Akteure entlang der Wertschöpfungskette beteiligt sind. Es lässt sich somit feststellen, dass das Thema Industrie

4.0 einen völlig neuen Blickwinkel auf den Bereich der Sicherheit wirft; s. (Kagermann et al. 2013, S. 50).

Grundsätzlich lässt sich das Thema Sicherheit in zwei Bereiche aufteilen. Zum einen muss das Thema der Betriebssicherheit (Safety) und zum anderen die Thematik der Angriffssicherheit (Security) berücksichtigt werden. Bei dem Thema Betriebssicherheit dürfen Maschinen, Produktionsanlagen, Werkstücke und Produkte zu keinem Zeitpunkt eine Gefahr für den Menschen darstellen. Die funktionale Sicherheit (Functional Safety) und eine hohe Zuverlässigkeit (Reliability) sind Voraussetzung für die Betriebssicherheit. Eine Sicherstellung dieser Bedingungen erweist sich als komplex, da hohe Anforderungen an verteilte IT-Architekturen in einer Produktion bestehen. Es wird erwartet, dass die Anlagen in der Produktion ohne Unterbrechung über mehrere Jahre hinweg laufen.

Der zweite Bereich ist das Thema Security und schützt Daten und Dienste in Systemen gegen Missbrauch, unbefugten Zugriff, Veränderungen und Zerstörungen. Das Ziel von Security ist die Erhöhung der Vertraulichkeit und der Integrität; s. (Zillmann 2016, S. 22).

Damit das Konzept der Industrie 4.0 realisiert und umgesetzt werden kann, müssen alle relevanten Risikobereiche der Sicherheit betrachtet werden. Zu diesen gehören Produktion und Fertigung mit ihren Embedded Systems, Big Data, Cloud-Computing und die mobilen Lösungen. Hieraus lässt sich erkennen, dass neue Formen der Zusammenarbeit entstehen und neue, zumindest angepasste IT-Sicherheitskonzepte erforderlich werden.

Mögliche Gefahren im Industrie 4.0-Bereich sind die Infektion mittels Schadsoftware über das Internet oder Intranet und über externe Hardware. Neben den gezielten Angriffen müssen sich Unternehmen auch gegen unabsichtlich herbeigeführte Probleme wappnen, wie beispielsweise dem menschlichen Fehlverhalten. Außerdem stellen die Fernwartungszugänge, die Nutzung mobiler Endgeräte und Distributed-Denial-of-Service (DDoS) weitere Gefahren innerhalb einer modernen Produktionsumgebung dar; s. (Huber 2016a, S. 65ff.). DDoS-Angriffe sind ein schwer abzuschätzendes Risiko für jede extern erreichbare Schnittstelle. Es werden viele Anfragen an einen Empfänger gestellt, so dass eine Überlastung der Server und Netzwerke vorliegt und legitime Anfragen nicht mehr bearbeitet werden können; s. (Bundesamt für Sicherheit in der Informationstechnik 2016, S. 28).

Ein weiteres Bedrohungsszenario geht einher mit der angestrebten Standardisierung der Anlagen und Maschinen. Die Informationen über Systeme und deren Spezifikationen werden im Internet weltweit verfügbar sein und sind folglich leicht manipulierbar; s. (Huber 2016a, S. 66).

Der Schutz von personenbezogenen Daten ist ein weiterer wichtiger Aspekt. Aufgrund der starken Vernetzung der Prozesse und der Digitalisierung muss dem Thema

große Aufmerksamkeit geschenkt werden (Huber 2016a, S. 68).

Der Ansatz BYOD (Bring your own Device) bringt ebenfalls viele Bedrohungen mit sich. Im Jahr 2013 nutzten circa 23 Prozent der Mitarbeiter in deutschen Unternehmen ihr eigenes mobiles Endgerät für Unternehmenszwecke. Dieser Trend verursacht eine Kostenersparnis für die Unternehmen, da die Anschaffungskosten für Mobile Devices entfallen. Die Mitarbeiter sind bereits mit ihrem mobilen Endgerät vertraut, wodurch die Mitarbeiterzufriedenheit gesteigert werden kann. Nicht zuletzt bringt die ortsungebundene Arbeit mehr Flexibilität und Produktivität in den Arbeitsalltag der Mitarbeiter. Neben den Vorteilen lassen sich auch viele Gefahren aufzeigen. Eine übergreifende Kontrolle kann nicht erfolgen und basiert lediglich auf dem entgegengebrachten Vertrauen an die Mitarbeiter, welche sich an die vereinbarten Richtlinien halten müssen. Des Weiteren sind auf jedem mobilen Device persönliche und viele unternehmensbezogene Daten gespeichert; s. (Stoitsi 2015). Mit Hilfe der zentralen Verwaltungsplattform MDM (Mobile Device Management) und EMM (Enterprise Mobile Management) können mobile Geräte mit zusätzlichem Tracking und Reporting in die Unternehmensinfrastruktur miteingebunden werden. Das MDM-System muss standardmäßig die Datenverschlüsselung von Datenträgern vornehmen, um somit die Sicherheit von Endgeräten gewährleisten zu können; s. (Huber 2016a, S. 65). Zur Unterstützung der Unternehmen werden bereits unternehmensbezogene Software und Applikationen für Mobile Devices entwickelt; s. (Stoitsi 2015).

Damit die die Produktionsstabilität innerhalb einer Smart Factory gewährleistet werden kann, müssen die Automatisierungsnetzwerke segmentiert werden. Die Automatisierungsnetzwerke werden in kleine Zellen aufgeteilt, zwischen denen Firewalls die Kontrolle übernehmen; s. (Huber 2016a, S. 67). Es ist sinnvoll, Firewalls mit Deep Packet Inspection (DPI)-Fähigkeit bei betriebskritischen Geräten einzusetzen. Bei DPI werden spezielle Befehle oder Anfragen geprüft, ob diese durch die Firewall gelassen werden dürfen oder nicht. So können unzulässige Zugriffe auf Produktionsobjekte vermieden werden; s. (Jenker 2014). Grundsätzlich sollten jedoch die altbekannten Maßnahmen nicht vergessen werden, wie beispielsweise Kameras, Kartenlesegeräte, organisatorische Maßnahmen und Passwortkontrolle. Ebenfalls sind Positivlisten ein einfacher Schutz gegen unkontrollierte Zugriffe. Hier erfolgt die Festlegung, auf welchem Rechner welche Prozesse, Programme oder Operationen laufen dürfen. Ein weiterer Ansatz zur Erhöhung der Sicherheit, ist das sogenannte Security by Design. Es ist von Vorteil, bereits bei der Produktentwicklung gewisse Sicherheitsaspekte zu berücksichtigen. Oftmals ist es nicht ausreichend, die CPS-basierten Produktionssysteme nachträglich um Security-Funktionen anzureichern; s. (Huber 2016a, S. 67f.).

Mit Hilfe des Security by Design-Ansatzes, wurde das plattformunabhängige und weltweit anerkannte Kommunikationsprotokoll OPC UA (Unified Architecture) entwickelt (siehe Kapitel „Standardisierung“). Dadurch entsteht eine sichere Vernetzung der Produktionsprozesse und stellt einen zentralen Baustein auf dem Weg zur Industrie 4.0 dar; s. (Bundesamt für Sicherheit in der Informationstechnik 2016, S. 52f.).

Viele Ansätze zur Einführung von geeigneten IT-Sicherheitsmaßnahmen kommen vom Bundesamt für Sicherheit und Informationstechnik (BSI). Das BSI beschäftigt sich mit sehr vielen IT-Sicherheitsthemen, von denen die meisten jedoch noch in der Entwicklungsphase stecken. Grundlegend können der IT-Grundschutz, das ICS-Security-Kompendium (Industrial Control Systems) und LARIS ICS als Hauptthemen für das BSI genannt werden. Der IT-Grundschutz ist der meistgenutzte Standard für Informationssicherheit in Deutschland und befindet sich gerade in einem Modernisierungsprozess. Das ICS-Security-Kompendium enthält die Grundlagen der IT-Sicherheit und die relevanten Normen und Standards. Des Weiteren beinhaltet es Best Practices und Empfehlungen zum Thema Cyber-Security für Anlagenbetreiber. Die Software LARIS ICS ist dagegen wie ein Tool aufgebaut, welches einen Fragebogen sowie Empfehlungen für weitere Maßnahmen umfasst. Das Tool trägt dazu bei, ein umfassendes IT-Sicherheitsmanagement in Unternehmen erarbeiten zu können; s. (Schonscheck 2015a).

Die IT-Sicherheitserfahrung in Industriebetrieben kann aktuell als eher gering eingeschätzt werden. Aus diesem Grund muss das BSI weitere Entwicklungs- und Optimierungsphasen durchlaufen, um den Unternehmen den Einstieg in die industrielle IT-Sicherheit so einfach wie

möglich zu gestalten; s. (Schonscheck 2015b). Dennoch bleibt die IT-Sicherheit eine individuelle Aufgabe jedes einzelnen Unternehmens. Diese müssen sich zum Ziel setzen, ein System mit einer inhärenten IT-Sicherheit zu konzeptionieren. Festzuhalten ist aber, dass es noch viele offene Baustellen im Bereich der IT-Sicherheit gibt und viele Ansätze noch nicht reif für den produktiven Einsatz sind; s. (Huber 2016a, S. 68).

### 5.3 Verfügbarkeit der IT-Infrastruktur

Eine weitere Herausforderung im Industrie 4.0-Umfeld ist die Verfügbarkeit der IT-Infrastruktur. Durch die stärkere Verwendung von Software und vernetzten Maschinen und Anlagen, steigt die Abhängigkeit der Unternehmen von einer leistungsstarken, skalierbaren und verfügbaren IT-Infrastruktur; s. (Huber 2016b). Aus diesem Grund muss die IT eine moderne und virtualisierte IT-Landschaft schaffen. Mit Hilfe von Standardisierung und Konsolidierung der IT-Systeme kann dies ermöglicht werden. Des Weiteren muss eine Vernetzung aller Komponenten und Systeme im Unternehmen erfolgen sowie die Gewährleistung der Verfügbarkeit von konsistenten Daten. Um diesen Anforderungen gerecht zu werden, muss eine Auflösung der klassischen Automatisierungspyramide, s. die Abbildung 4 erfolgen; s. (Zillmann 2016, S. 20).

Aktuell wird die Kommunikation in Unternehmen durch das hierarchische System der Automatisierungspyramide widerspiegelt. Das Ziel der Automatisierungspyramide ist es, die Komplexität der industriellen Fertigung zu minimieren. Dies wird durch die Einteilung der Unternehmensprozesse in die einzelnen Ebenen erreicht. Die ver-

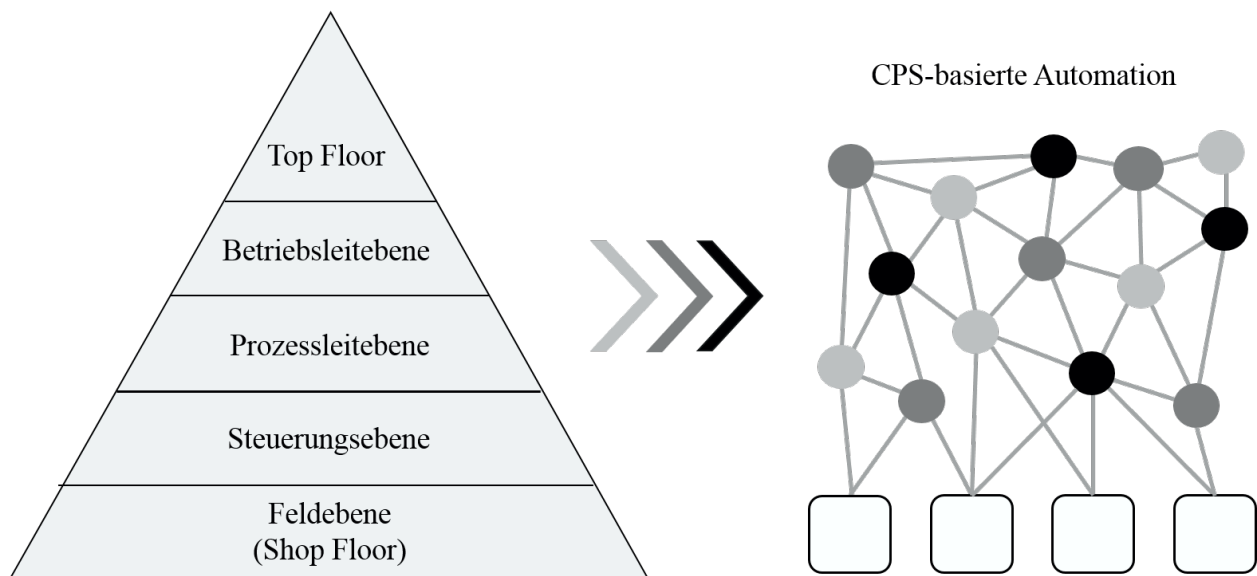


Abbildung 4: Traditionelle Automatisierungs-Pyramide und CPS-basierte Automatisierung (Eigene Darstellung)

schiedenen Ebenen werden jeweils durch unterschiedliche Systeme unterstützt; s. (Meudt et al. 2017, S. 7). Eine Unterstützung erfolgt in der Unternehmensleitebene mittels der Enterprise Resource Planning (ERP) Software und wird als Informationssystem in allen Bereichen eines Unternehmens für wirtschaftliche Tätigkeiten eingesetzt. Des Weiteren wird in der Betriebsleitebene die Software Manufacturing Execution System (MES) genutzt und ist mit den Aufgaben Produktionssteuerung, -lenkung und -kontrolle vertraut. Darunter befindet sich die Produktionsleitebene, in der die kurzfristige Produktionsplanung erfolgt. Dies geschieht mit Hilfe von Systemen, wie beispielsweise dem Supervisory Control and Data Acquisition (SCADA)-System. Während dessen befinden sich in der Prozessleitebene die Steuer- und Regelsysteme, wie die sogenannte speicherprogrammierbare Steuerung (SPS). In dieser Ebene werden die Signale von der Feldebene ausgewertet und weiterverarbeitet an die Feldebene zurückgegeben. Die Feldebene ist mit Sensoren und Aktoren ausgestattet und liefert produktionsrelevante Daten an die übergeordneten Ebenen; s. (Heinrich 2017, S. 5).

Heutzutage erfolgt die Kommunikation noch hierarchisch über wenige Schnittstellen zwischen den einzelnen Ebenen. Aus diesem Grund erfolgt die Betrachtungsweise und die Steuerung der Ebenen in der Regel noch isoliert; s. (Bauernhansl et al. 2014, S. 544). Der Schritt zu einer übergreifenden Vernetzung ist derzeit im Bereich der Automatisierungstechnik kaum vollzogen. Bereits heute liegen sehr große Mengen an Informationen über Produktionsmittel, Produkte und Prozesse in den IT-Systemen der Unternehmen vor. Ebenfalls existieren in Unternehmen meist eine Vielzahl an IT-Systemen und Maschinen nebeneinander, welche nur eine bedingte Vernetzung aufweisen oder welche aufgrund unterschiedlicher Semantik keine Datenmigration erlauben; s. (Bauernhansl et al. 2014, S. 39). Durch die Isolation der Informationsquellen werden Medienbrüche verursacht. Es lässt sich feststellen, dass der Nutzen der Informationsinterpretation in einer konkreten Anwendung erst dann erfolgen kann, wenn unterschiedliche Subsysteme gemeinsam integriert werden; s. (Bauernhansl et al. 2014, S. 79). Zudem kommt das Problem, dass die Systeme, Maschinen und Anlagen meist von unterschiedlichen Herstellern bezogen werden und ein unterschiedliches Alter aufweisen. Aus diesen Gründen ist das Umrüsten der Automatisierungssoftware sehr aufwändig, um die benötigte Kompatibilität herzustellen. Eine weitere Herausforderung ist es, den Datenfluss zu angrenzenden internen und externen Bereichen herzustellen; s. (Schröder 2016, S. 11). Ein ideales Ergebnis wird erst erreicht, wenn die Auflösung der starren hierarchischen Automatisierungspyramide erfolgt. Dies impliziert jedoch nicht das vollständige Verschwinden der Systeme auf den unterschiedlichen Ebenen. Der Fokus liegt hier vielmehr auf einer nahtlosen Verbindung zwischen den Ebenen zu einer Netzwerkstruktur. Damit dieses Ziel erreicht werden kann, müssen zeit- und finanztechnische Bemühungen der Integration erbracht werden; s. (Meudt et al.

2017, S. 7). Bei dem Thema Integration werden die flexible An- und Abkoppelbarkeit der einzelnen Komponenten, die programmieretechnischen Aufwände sowie die IT-Sicherheit der verteilten Informationen betrachtet. Für die anwendungsspezifische Verteilung von Informationen werden sogenannte Middleware-Systeme eingesetzt und führen Informationen aus verschiedene Quellsysteme zusammen. Diese werden über einheitliche Schnittstellen in technologieunabhängiger Form für Interpretationssysteme zur Verfügung gestellt. Wird der initiale Aufbau einer solchen Kommunikationsarchitektur aus Sicht der zu tätigen Investitionen betrachtet, so stellt dies den größten Aufwand dar. Ist die grundlegende Architektur realisiert, so können an die bestehende Infrastruktur weitere Datenquellen oder Interpretationssysteme mit überschaubarem Aufwand angebunden werden. Andererseits stellen bei den technischen Umsetzungen von Anwendungen die Identifikation der auszutauschenden Daten, die Umsetzung der Anbindung von Datenquellen und die Interpretationssysteme derzeit die größten Aufwände dar. In aktuellen Projekten, wie beispielsweise dem Forschungsprojekt CyProS (Cyber-Physische Produktionssysteme), wird an Komponenten geforscht, die die Umsetzbarkeit solcher Architekturen entscheidend beschleunigen; s. (Bauernhansl et al. 2014, S. 79ff.).

Grundsätzlich bedarf es nationale und internationale Abstimmungen bezüglich Standards für die Kommunikation zwischen den einzelnen Hardware- und Software-Applikationen. Daher muss den Unternehmen von Eigenentwicklungen abgeraten werden, da für eine verfügbare und sichere IT-Struktur internationale Standards benötigt werden; s. (Meudt et al. 2017, S. 7). Damit die Auflösung der Automatisierungspyramide und somit ein Aufbau einer verfügbaren IT-Infrastruktur gelingt, bedarf es noch einige Zeit.

#### 5.4 Verfügbarkeit von schnellem Internet

Ein ebenso wichtiges Thema ist die Verfügbarkeit von schnellem Internet. Bezüglich des Ausbaus und der Verfügbarkeit der Breitbandnetze, befindet sich Deutschland im internationalen Vergleich auf den hinteren Plätzen. Doch gerade im Zeichen der Industrie 4.0 und der Digitalisierung ist der Zustand „Always on“ ein absolutes Muss; s. (Huber 2016b). Ein flächendeckender Einsatz von CPS setzt eine Infrastruktur voraus, mit der ein höherer und qualitativ hochwertigerer Austausch der Daten ermöglicht wird. Damit eine geringe Latenzzeit, eine hohe Ausfallsicherheit, eine hohe Qualität und ein umfassendes Breitbandnetz umgesetzt werden können, muss ein Ausbau der bestehenden Kommunikationsnetze erfolgen; s. (Kagermann et al. 2013, S. 49f.). Als Breitbandverbindung definiert die International Telecommunication Unit (ITU) bereits eine Übertragungsrate von zwei Megabit pro Sekunde. Diese festgelegte Übertragungsgeschwindigkeit ist nicht ausreichend, um unternehmensübergreifend und internetbasiert eine Produktion zu organisieren oder nachgelagerte Dienstleistungen zu definieren. Aus diesem Grund sind stabile Hochleistungsübertragungswege über Glasfaserkabel notwendig.



Ein weiteres Argument, den Glasfaserausbau voranzutreiben, ist, dass dieser nach heutigem Stand der Technik alternativlos ist. Ohne Glasfaserkabeln kann eine Smart Factory nicht realisiert werden und macht den Ausbau der Breitbandverbindung somit unabdingbar; s. (Schröder 2016, S. 1).

Die Bundesregierung erkennt den dringenden Bedarf an schnellem Internet und reagiert mit Förderungsprogrammen, Gesetzen und Strategien. Das Ziel der Bundesregierung ist, dass alle Haushalte bis zum Jahr 2018 einen Zugang zu schnellem Internet mit einer Übertragungsrate von mindestens 50 Megabit pro Sekunde haben. Des Weiteren setzt sich die Regierung dafür ein, bis 2025 ein flächendeckendes Breitbandnetz zu errichten, um auch Gewerbe-, Industriegebiete und Häfen an das Glasfasernetz anzuschließen. Die öffentlich zugänglichen Flächen der Gewerbe- und Industriegebiete werden zugleich mit kostenlosem WLAN (Wireless Local Area Network) ausgestattet. Das Gesetz „Gesetz zur Erleichterung des Ausbaus digitaler Hochgeschwindigkeitsnetze“ (DigiNetz-Gesetz) wurde 2016 von der deutschen Bundesregierung verabschiedet und fördert den Breitbandausbau in noch nicht erschlossenen Gebieten. Mittels dem DigiNetz-Gesetz werden bei einer Neuerschließung von Wohn- und Gewerbegebieten die Glasfaserleitungen künftig immer mitverlegt.

Auch in der Mobilfunktechnik gibt es einige Neuerungen. Im Jahr 2015 stellte Deutschland bereits Frequenzen im 700 Megahertz-Bereich für den Mobilfunk bereit und etablierte sich so zum europäischen Vorreiter. Doch aktuell wird eine neue Generation des Mobilfunks, namens 5G-Netzstandard, entwickelt. Mit dieser neuen Technologie lassen sich Datenmengen zuverlässig mit bis zu 20 Gigabit pro Sekunde übertragen. Aus diesen Gründen entwickelte die Bundesregierung die „5G-Strategie Deutschland“. Die 5. Generation des Mobilfunks wird damit zur digitalen Schlüsseltechnologie im Zeitalter der Vernetzung und ermöglicht eine Vielzahl an neuen Geschäftsmodellen, wie beispielsweise dem autonomen und vernetzten Fahren und der Industrie 4.0. Die 5G-Technik wird laut Bundesregierung bis 2020 marktreif sein; s. (Die Bundesregierung 2017).

## 5.5 Komplexe Systeme

Das nächste Problem, welches mit der Industrie 4.0 einhergeht, ist die zunehmende Komplexität innerhalb einer Smart Factory. Aufgrund der technologischen Entwicklungen in den letzten Jahrzehnten, ist die Komplexität der Produkte und Systeme bereits sehr stark gewachsen. Dies hat zur Folge, dass auch die Komplexität der industriellen Prozesse der Entwicklung und Fertigung stark ansteigen. Die Entwicklung und Produktion der Güter erscheint somit immer weniger beherrschbar; s. (Sendler 2013, S. 16).

Dass ein zu hohes Maß an IT und Automatisierung keine Allheilmittel sind, hat sich bereits in der Vergangenheit

herausgestellt. Die neu eingeführten Technologien können nicht mehr ausreichend beherrscht werden und die Unternehmen werden zunehmend unflexibel; s. (Huber 2016b). Die Rede ist vom Computer Integrated Manufacturing Ansatz (CIM Ansatz), welcher in den 80er Jahren verfolgt wurde. Beginnend von der Planung bis hin zur Fertigung sollte eine Vollautomatisierung erfolgen. Der Ansatz scheiterte, weil zum damaligen Zeitpunkt die erforderlichen Technologien nicht vorhanden oder zu unerschwinglichen Preisen erhältlich waren. Es wurde eine überzüchtete und überteuerte Produktion geschaffen, welche nur noch schwer beherrschbar war. Aufgrund der Komplexitätsprobleme und mangelnder Wirtschaftlichkeit wurde der Ansatz zu dieser Zeit nicht weiterverfolgt; s. (Bauernhansl et al. 2014, S. 85f.).

Im Laufe der Zeit verbesserte sich die Leistungsfähigkeit der Technologien und ebenso die Fähigkeiten der Entwickler. Zum heutigen Zeitpunkt haben genau hier die Ingenieure in der Mitte Europas eine gute Position, da sie die Verbindung von Embedded Software mit industriellen Produkten verschiedenster Arten sehr gut beherrschen. Manche Unternehmen schaffen die Verbindung von Software und Elektronik mit Mechanik besser als andere und sichern sich somit einen Platz an der Spitze des Weltmarkts. Dieser Vorsprung beruht auf den Erfahrungen, die sich über die letzten Jahrzehnte in der Industrie angesammelt haben. Es wird aber deutlich, dass der Vorsprung nicht langfristig gesichert ist. Die Voraussetzungen sind jedoch sehr gut, diesen zu halten oder sogar auszubauen. Der Industrie muss die Anpassung der Fähigkeiten, Methoden, Prozessen und Geschäftsmodellen an den Herausforderungen vernetzter, softwaregesteuerter Systeme gelingen. Ist dieser Prozess der Veränderung erfolgreich, so kann ein globaler Wettbewerbsvorteil entstehen. Die europäischen Produkte und Systeme würden neben der hohen Qualität, Langlebigkeit und Umweltverträglichkeit ebenfalls wegen Sicherheit punkten, da eine zukünftige Zusammenarbeit von vielen Fachdisziplinen bei Entwicklung und Bau der Systeme erfolgt. Durch diese Kooperation können komplexe Systeme problemlos ablaufen und Prozesse so einfach wie möglich gestaltet werden. Zudem lernen die Ingenieure die Komplexität einer Smart Factory zu steuern. Grundlegend dafür ist das Verständnis des Zusammenspiels der verschiedenen Disziplinen innerhalb einer modernen Produktionsumgebung. Es muss den Ingenieuren gelingen, die einfache Bedienung der heutigen mobilen Endgeräte auf die Prozesse und Produkte der industriellen Entwicklung und Produktion zu übertragen. Dieser Vorgang wird oftmals als *Simplicity* bezeichnet; s. (Sendler 2013, S. 16f.).

## 5.6 Organisatorische Risiken

Im Allgemeinen reduziert sich der Bereich der Smart Factory fast ausschließlich auf die Einführung neuer Technologien. Dennoch müssen organisatorische Risiken abgewogen werden. Wird dies nicht getan, besteht die Gefahr, dass die durch die Industrie 4.0 herbeigehenden

den Veränderungen zu erheblichen Problemen, fallendem Potential und Verzögerungen in der Implementierung von Industrie 4.0 führen; s. (Huber 2016b).

Die Aufbauorganisation eines Unternehmens spielt besonders in seiner höchsten Hierarchieebene eine signifikante Rolle. Hier muss das Management eine klare Strategie und einen Plan zur Digitalisierung definieren, um das Verständnis für die IT und die Prozesse aufzuzeigen. Organisationsmodelle müssen außerdem Kooperationen ebenso wie die Kommunikation zwischen allen Teilnehmern (Menschen und Maschinen) ohne Hierarchiehürden ermöglichen. Um die Mitarbeiter entsprechend auf die neuen Aufgaben in einer Smart Factory einstellen zu können, muss das Management ein Change-Team bzw. eine Change-Abteilung aufstellen – in der Realität wird dies oftmals nicht umgesetzt. Dies hat eine Art Neophobie zur Folge, wobei die neuen Aufgabenbereiche zunächst sehr kritisch betrachtet werden. Außerdem ist häufig eine sinkende Zufriedenheit der Mitarbeiter zu beobachten. Aus diesem Grund muss das Change-Team dafür sorgen, dass Kompetenzen ebenso wie die Qualifizierung der Angestellten des Unternehmens gegeben ist; s. (Huber 2016b).

Die neuen Technologien von Industrie 4.0 ändern sowohl die Arbeitsabläufe und -anforderungen der Ausbildungsplätze als auch das Berufsbild der einzelnen Tätigkeiten als solches. In der Produktion wird weitaus mehr Fachpersonal benötigt, gleichzeitig sinkt der Anteil der Mitarbeiter mit geringeren Qualifikationen, was auf die Digitalisierung zurückzuführen ist. Aktuell haben Unternehmen das Problem, dass viele ihrer Angestellten nicht in der Lage sind die neuen Technologien anzuwenden, da sie die neuen Prozesse nicht verstehen. „Lifelong learning“, Veränderungen in der Arbeitsorganisation, Teamzusammensetzung und Wissensmanagement werden auch noch in der Zukunft eine signifikante Rolle spielen; s. (Huber 2016b).

## 5.7 Wirtschaftliche Risiken

Zusätzlich zu den technischen und organisatorischen Risiken müssen nun die wirtschaftlichen Risiken abgewogen werden. Dazu wurden verschiedene Herausforderungen berücksichtigt. Die erste Hürde ist der fehlende Wille zu investieren – obwohl Investitionen in Industrie 4.0 im Moment relativ kostengünstig sind. Das größte Problem stellt jedoch die fehlende Transparenz des wirtschaftlichen Mehrwerts dar. So müssen Unternehmen in der Maschinenbau- und Anlagenindustrie schon heute zum Zeitpunkt der Investition die Anforderungen erwägen, die Produkte in der Zukunft haben werden. Grundsätzlich lässt sich feststellen, dass die Unternehmen die Relevanz von Investitionen in Industrie 4.0 in allen Sektoren schon lange erkannt und wahrgenommen haben. Unternehmen müssen die notwendigen Mittel für Investitionen und Ressourcen bereitstellen, damit eine Smart Factory umgesetzt werden kann. Studien zeigen, dass in Deutschland

jährlich 40 Milliarden Euro in Industry 4.0-Anwendungen investiert werden; s. (Huber 2016b).

Ein weiteres Risiko ist der Verlust von Arbeitsplätzen innerhalb einer Smart Factory, da einfache Produktionsarbeiten durch die Digitalisierung ersetzt werden. Ziel der Einführung einer Smart Factory sollte es jedoch sein, die Automatisierung zu erhöhen und die Kosten zu senken und nicht, die Zahl der Arbeitsplätze zu reduzieren (Huber 2016b). Die Einführung einer Smart Factory verlangt den flexiblen Einsatz von Personal. Es gibt bereits starke Schwankungen im Personalbedarf. In Zukunft werden diese Schwankungen noch größer sein. Auch die Anforderungen an den Produktionsmitarbeiter in Bezug auf die Qualifikation ändern sich; s. (Huber 2016b). Akademischen Berufe wie Mathematiker, Informatiker, Berufe im Bereich der Naturwissenschaften und Technik müssen ebenso wie die nicht-akademischen Berufe der Mechatronik und Automatisierungstechnik weiter gefördert werden, um die hohe Nachfrage in der Wirtschaft zu befriedigen. Die Ausbilder müssen die Fähigkeiten der Auszubildenden in den Bereichen der Industrie 4.0-Technologien steigern. Durch interne Aus- und Weiterbildung wird sichergestellt, dass die Mitarbeiter in der Produktion auf dem neuesten Stand der Technik sind. Die Einführung der Smart Factory ermutigt die Mitarbeiter, sich selbst zu organisieren und damit mehr Eigenverantwortung zu übernehmen; s. (Huber 2016b).

Folglich kann der Verlust einiger Geschäftsbereiche als ein wirtschaftliches Risiko bezeichnet werden. Die neuen Marktteilnehmer verdrängen etablierte Hersteller wie Buchhandlungen und Verlage während sogenannte Start-ups den Herstellern neue Wege öffnen. Zudem ist die mangelnde Unterstützung der Forschung ein großes Hindernis. Deutschland investiert gut in Forschung und Finanzierung, jedoch investieren Länder wie China und Japan durch ihre höhere Investitionsbereitschaft besser im selben Bereich; s. (Huber 2016b).

Die festgestellten wirtschaftlichen Risiken deuten darauf hin, dass ohne Berücksichtigung dieser Fragen keine wettbewerbsfähige und vernetzte Produktionsumgebung entstehen kann.

## 6. Fazit

Dieses Dokument gibt dem Leser einen strukturierten Überblick über die Risiken, die in einer Intelligenten Fabrik auftreten können. Es werden verschiedene Risiken in den Bereichen Technologie, Organisation und Wirtschaft betrachtet. Befasst man sich etwas genauer mit den verschiedenen Risikobereichen wird allerdings deutlich, dass in vielen Bereichen noch Handlungsbedarf besteht. Unternehmen sind vor allem in den technischen Bereichen auf (Bundes-)Hilfen angewiesen, um die Digitalisierung der Wirtschaft voranzutreiben.

Die Implementierung von Industrie 4.0 gelingt nicht einzig durch die Hilfe der Bundesregierung, sondern erfor-

dert auch die Unterstützung der Unternehmen. Es ist notwendig, dass die Unternehmen und ihr Management für das Thema Smart Factory und Digitalisierung sensibilisiert werden und die Dringlichkeit einer Veränderung in der Produktion wahrnehmen. Wenn dies von den deutschen Unternehmen nicht berücksichtigt wird, wird sich diese Nichtbeachtung in den kommenden Jahren negativ auf die Wettbewerbsfähigkeit Deutschlands auswirken. Eine Anpassung der Produktionsmitarbeiter innerhalb der Smart Factory ist ebenfalls erforderlich.

Da die vierte industrielle Revolution erst am Anfang steht und noch die Möglichkeit besteht, die sich abzeichnenden Risiken klar zu erörtern, können so geeignete Gegenmaßnahmen ergriffen werden. Auf diese Weise können die Themen von Industry 4.0 erfolgreich umgesetzt und die sich abzeichnenden Potenziale genutzt werden.

Es konnten nicht alle Risiken einer vernetzten Produktionsumgebung angegangen werden. Ebenso wurden nicht alle Risiken detailliert und vollständig dargestellt. Es wurden ausgewählte Risiken beschrieben, die als relevant eingestuft wurden. Da dieses Thema sehr aktuell ist und sich noch in der Entwicklungsphase befindet, müssen die Untersuchungen parallel zum Entwicklungsfortschritt vorangetrieben werden. Aus diesem Grund sind weitere Untersuchungen in diesem Bereich notwendig und empfehlenswert, um einen vollständigen Überblick über die Intelligente Fabrik und ihren Status zu erhalten.

### Literaturverzeichnis

Bauer, Wilhelm et al. (2018): Künstliche Intelligenz in der Unternehmenspraxis – Studie zu Auswirkungen auf Dienstleistung und Produktion

Bauernhansl, Thomas; Hompel, Michael ten; Vogel-Heuser, Birgit (Hg.) (2014): Industrie 4.0 in Produktion, Automatisierung und Logistik. Anwendung, Technologien, Migration. Wiesbaden: Springer Vieweg (SpringerLink). Online verfügbar unter <http://dx.doi.org/10.1007/978-3-658-04682-8>.

Botthof, Alfons et al. (2015): Zukunft der Arbeit in Industrie 4.0

Bundesamt für Sicherheit in der Informationstechnik (2016): Die Lage der IT-Sicherheit in Deutschland 2016. Hg. v. Bundesamt für Sicherheit in der Informationstechnik. Online verfügbar unter [https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2016.pdf?\\_\\_blob=publicationFile&v=4](https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2016.pdf?__blob=publicationFile&v=4), zuletzt geprüft am 06.12.2020.

Bundesministerium für Wirtschaft und Energie: Dossier "Industrie 4.0". Digitale Transformation in der Industrie. Unter Mitarbeit von Bundesministerium für Wirtschaft und Energie. Hg. v. Bundesministerium für Wirtschaft und Energie. Online verfügbar unter <https://www.bmwi.de/Redaktion/DE/Dossier/industrie-40.html>, zuletzt geprüft am 07.12.2020.

Die Bundesregierung (2017): Von der digitalen zur Gigabit-Gesellschaft. Breitbandausbau. Hg. v. Die Bundesregierung. Online verfügbar unter <https://www.bundesregierung.de/Content/DE/Artikel/2017/08/2017-08-30-breitbandausbau.html>, zuletzt geprüft am 06.11.2020.

Bischoff, Jürgen (Hg.) (2015): Studie "Erschließung der Potenziale der Anwendung von "Industrie 4.0" im Mittelstand". Studie im Auftrag des Bundesministeriums für Wirtschaft und Energie (BMWi). Unter Mitarbeit von agiplan GmbH, Fraunhofer IML und ZENIT GmbH. Online verfügbar unter [http://www.zenit.de/fileadmin/Downloads/Studie\\_im\\_Auftrag\\_des\\_BMWi\\_Industrie\\_4.0\\_2015\\_agiplan\\_fraunhofer\\_iml\\_zenit\\_Langfassung.pdf](http://www.zenit.de/fileadmin/Downloads/Studie_im_Auftrag_des_BMWi_Industrie_4.0_2015_agiplan_fraunhofer_iml_zenit_Langfassung.pdf), zuletzt geprüft am 25.11.2020.

Hankel, Martin; Rexroth, Bosch (2015): Industrie 4.0: Das Referenzarchitekturmodell Industrie 4.0 (RAMI 4.0). Hg. v. ZVEI - Zentralverband Elektrotechnik- und Elektronikindustrie e. V. Online verfügbar unter [https://www.zvei.org/fileadmin/user\\_upload/Themen/Industrie\\_4.0/Das\\_Referenzarchitekturmodell\\_RAMI\\_4.0\\_und\\_die\\_Industrie\\_4.0-Komponente/pdf/ZVEI-Faktenblatt-Industrie4\\_0-RAMI-4\\_0.pdf](https://www.zvei.org/fileadmin/user_upload/Themen/Industrie_4.0/Das_Referenzarchitekturmodell_RAMI_4.0_und_die_Industrie_4.0-Komponente/pdf/ZVEI-Faktenblatt-Industrie4_0-RAMI-4_0.pdf), zuletzt geprüft am 02.12.2020.

Heinrich, Berthold (2017): Grundlagen zur Automatisierung: Springer Fachmedien (SpringerLink). Online verfügbar unter 10.1007/978-3-658-17582-5\_1, zuletzt geprüft am 06.12.2020.

Huber, Walter (2016a): Industrie 4.0 in der Automobilproduktion. Ein Praxisbuch. Wiesbaden: Springer Vieweg.

Huber, Walter (2016b): Industrie 4.0 und die Risiken. Online verfügbar unter <https://www.computerwoche.de/a/industrie-4-0-und-die-risiken,3324008>, zuletzt geprüft am 15.11.2020.

Jasperneite, Jürgen (2012): Was hinter Begriffen wie Industrie 4.0 steckt. Internet und Automation. Online verfügbar unter <http://www.computer-automation.de/steuerungsebene/steuern-regeln/artikel/93559/>, zuletzt geprüft am 30.11.2020.

Jenker, Martin (2014): Cyber-Sicherheit für Steuer- und Regelungsnetzwerke. Cell-Zone-Site-Konzept. Hg. v. SPS Magazin. Online verfügbar unter [http://www.spsmagazin.de/?inc=artikel/article\\_show&nr=87493](http://www.spsmagazin.de/?inc=artikel/article_show&nr=87493), zuletzt geprüft am 06.12.2020.

Kagermann, Henning; Wahlster, Wolfgang; Helbig, Johannes (2013): Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0. Deutschlands Zukunft als Produktionsstandort sichern - Abschlussbericht des Arbeitskreises Industrie 4.0. Online verfügbar unter [https://www.bmbf.de/files/Umsetzungsempfehlungen\\_Industrie4\\_0.pdf](https://www.bmbf.de/files/Umsetzungsempfehlungen_Industrie4_0.pdf), zuletzt geprüft am 30.11.2020.

Kelkar, Oliver; Heger, Roland; Dao, Dan-Khanh (2014): Studie Industrie 4.0 - Eine Standortbestimmung der Automobil- und Fertigungsindustrie. Hg. v.

Mieschke Hofmann und Partner (MHP). Online verfügbar unter [https://www.mhp.com/fileadmin/mhp.de/assets/studien/MHP-Studie\\_Industrie4.0\\_V1.0.pdf](https://www.mhp.com/fileadmin/mhp.de/assets/studien/MHP-Studie_Industrie4.0_V1.0.pdf), zuletzt geprüft am 15.11.2020.

Krüger, Jörg et al. (2019): WGP Standpunkt, KI in der Produktion – Künstliche Intelligenz erschließen für Unternehmen

Mauerer, Jürgen (2016): Was ist was bei Industrie 4.0? Begriffe rund um die vernetzte Produktion. Online verfügbar unter <http://www.computerwoche.de/a/was-ist-was-bei-industrie-4-0,3313199>, zuletzt geprüft am 12.11.2020.

Meudt, Tobias; Pohl, Malte; Metternich, Joachim (2017): Die Automatisierungspyramide - Ein Literaturüberblick. Hg. v. Technische Universität Darmstadt. Online verfügbar unter <http://tuprints.ulb.tu-darmstadt.de/6298/1/2017%20-%20Die%20Automatisierungspyramide%20-%20Ein%20Literatur%C3%BCberblick-2.pdf>, zuletzt geprüft am 06.12.2020.

Münzl, Gerald; Pauly, Michael; Reti, Martin (Hg.) (2015): Cloud Computing als neue Herausforderung für Management und IT. Berlin: Springer Vieweg (essentials).

Plattform Industrie 4.0 (2015): Umsetzungsstrategie Industrie 4.0. Ergebnisbericht der Plattform Industrie 4.0. Online verfügbar unter <https://www.bitkom.org/Publikationen/2015/Leitfaden/Umsetzungsstrategie-Industrie-40/150410-Umsetzungsstrategie-0.pdf>, zuletzt geprüft am 12.11.2020.

Reinhart, Gunther et al.: Industrieroboter: Planung - Integration - Trends Ein Leitfaden für KMU, Würzburg 2018.

Ruskowski, Martin (2020): Von Industrie 4.0 zur KI in Service und Produktion. Vortrag. 16. Empolis Executive Forum 2019 in Berlin: 19.11.2020.

Schewe, Frank (2016): Was hinter der Referenzarchitektur RAMI 4.0 steckt. Industrie 4.0. Hg. v. computer-automation.de. Online verfügbar unter <http://www.computer-automation.de/steuerungsebene/steuern-regeln/artikel/129204/>, zuletzt geprüft am 06.12.2020.

Schonscheck, Oliver (2015a): Security-Konzept im Praxis-Check. IT-Sicherheit & Industrie 4.0. Spezielle Sicherheitskonzepte sind in Arbeit. Hg. v. Computerwoche. Online verfügbar unter <https://www.computerwoche.de/a/security-konzepte-im-praxis-check,3211303,2>, zuletzt geprüft am 06.12.2020.

Schonscheck, Oliver (2015b): Security-Konzepte im Praxis-Check. IT-Sicherheit & Industrie 4.0. IT-Sicherheitskonzepte noch nicht reif für die Industrie. Hg. v. Computerwoche. Online verfügbar unter <https://www.computerwoche.de/a/security-konzepte-im-praxis-check,3211303,3>, zuletzt geprüft am 06.12.2020.

Schröder, Christian (2016): Herausforderungen von Industrie 4.0 für den Mittelstand. Hg. v. Friedrich-Ebert-Stiftung. Online verfügbar unter

<http://library.fes.de/pdf-files/wiso/12277.pdf>, zuletzt geprüft am 06.12.2020.

Sendler, Ulrich (Hg.) (2013): Industrie 4.0. Beherrschung der industriellen Komplexität mit SysLM. Berlin, Heidelberg, s.l.: Springer Berlin Heidelberg (Xpert.press).

Stoitsi, Maria (2015): Die Vor- und Nachteile von ByoD. Enterprise Mobility Management. Hg. v. Computerwoche. Online verfügbar unter <https://www.computerwoche.de/a/die-vor-und-nachteile-von-byod,3213991>, zuletzt geprüft am 06.12.2020.

Termer, Frank (2019): Künstliche Intelligenz und ERP. Online verfügbar unter <https://www.bitkom.org/Bitkom/Publikationen/Kuenstliche-Intelligenz-und-ERP>, zuletzt geprüft am 10.11.2020

VDI/VDE-Gesellschaft (2013): Thesen und Handlungsfelder. Cyber-Physical Systems: Chancen und Nutzen aus Sicht der Automation. Online verfügbar unter [https://www.vdi.de/uploads/media/Stellungnahme\\_Cyber-Physical\\_Systems.pdf](https://www.vdi.de/uploads/media/Stellungnahme_Cyber-Physical_Systems.pdf), zuletzt geprüft am 30.11.2020.

Zillmann, Mario (2016): Smart Factory - Wie die Digitalisierung Fabriken verändert. Transformation von der Werkshalle bis zur Unternehmensleitung. Hg. v. Lünen-donk GmbH in Zusammenarbeit mit T-Systems und Produktion. Online verfügbar unter <http://luegendonk-shop.de/Luegendonk-Publikationen/Luegendonk-Whitepaper-Smart-Factory-Wie-die-Digitalisierung-Fabriken-veraendert.html>.

# Verbesserung des Customer Services durch Natural Language Processing

Elis Bardhi

Technische Hochschule  
Mittelhessen

Fachbereich MND  
Wilhelm-Leuschner-Straße 13  
61169 Friedberg  
[elis.bardhi@mnd.thm.de](mailto:elis.bardhi@mnd.thm.de)

Prof. Dr. Harald Ritz

Technische Hochschule  
Mittelhessen

Fachbereich MNI  
Wiesenstraße 14  
35390 Gießen  
[harald.ritz@mni.thm.de](mailto:harald.ritz@mni.thm.de)

Tim-Niklas Förster

advades GmbH

SAP Consultant  
Kaiserstraße 7  
35745 Herborn  
[t.foerster@advades.com](mailto:t.foerster@advades.com)

## Kategorie

Bachelorarbeit

## Schlüsselwörter

Natural Language Processing, Natural Language Understanding, Natural Language Generation, Customer Service, Customer Service Automation, Ticket System, Customer Self Service, Natural Language IVR, Chatbot.

## Zusammenfassung

Mit dieser Bachelorarbeit wird dem Leser die Verbesserung des Customer Services durch den Einsatz der Natural Language Processing dargestellt. Mit der Entwicklung der NLP ist auch ihr Einfluss auf die Technologien wie Interactive Voice Response oder Customer Relationship Management gestiegen, die vom Customer Service benutzt werden. Durch deren Einsatz wurden viele Routineaufgaben automatisiert, sodass die Arbeit der Mitarbeiter des Customer Services effizienter geworden ist. Auch Prozesse, die aufgrund der großen Datenmenge eine lange Ausführungszeit hatten, werden aufgrund von NLP bzw. künstlicher Intelligenz in wenigen Sekunden erledigt.

Als erstes wird dem Leser eine Einführung in das Thema „Natural Language Processing“ gegeben. Es werden die Teilbereiche der NLP, und zwar „Natural Language Understanding“ und „Natural Language Generation“, sowie die Komponenten der NLP erklärt und aufgelistet. Darüber hinaus erfährt der Leser etwas über zentrale Anwendungsbereiche der NLP in den heutigen Technologien.

Als zweites wird eine Einführung in das Thema „Customer Service“ gegeben, unter anderem wie Customer Service funktioniert und welchen Einfluss dieser auf die Prozesse des Unternehmens hat. Dazu werden die Arten des Customer Services u.a. „Telefon“, „E-Mail“, „Live Chat“ sowie „Social Media Support“ aufgelistet und beschrieben. Die Teilbereiche des Customer Services, und zwar „Pre-Sales-“ und „After-Sales-Services“ und ihre Strategien sowie Vorteile, werden beschrieben.

Weiterhin wird die Automatisierung des Customer Services beleuchtet. Hier wird der Leser über die verschiedenen Anwendungen des Customer Services informiert, die mit Hilfe technologischer Fortschritte sowie der NLP optimiert wurden. Zu Beginn werden die Einsatzmöglichkeiten zur Verbesserung der Anfrageverwaltung („Ticket System“) sowie zur Kundenselbstbedienung („Customer Self Service“) beschrieben. Darüber hinaus werden die Vorteile der Automatisierung des Customer Services erläutert. Folgend erfährt der Leser etwas über die Verbesserung des Prozessablaufs dieser Applikationen sowie die Erfüllung der Kunden-erwartungen und -präferenzen durch NLP anhand von Beispielen.

Als Haupttechnologie zur Darstellung der Verbesserung des Customer Services durch NLP werden dem Leser die Arten und Funktionen des Chatbots dargestellt. Darüber hinaus lernt der Leser über die Einsatzvorteile davon in einem Unternehmen sowie die verschiedene Anwendungsbereiche. Für ein besseres Verständnis werden anhand der „SAP Conversational AI“-Plattform, deren Komponenten zum Logikaufbau sowie ihre Eigenschaften aufgelistet und erklärt.

Ein bekanntes Beispiel von Chatbots und ihrem Einfluss auf Customer Service stellt „AdmiralDirekt.de“ dar. Das Kfz-Versicherungsunternehmen verfügt über einen virtuellen Assistent, der u.a. Schadensmeldungen annimmt, Fahrzeuge für eine neue Versicherung anmeldet und Zulassungsunterlagen zusendet. Zusätzlich kann der Bot komplexe Anfragen verstehen und selbständig beantworten. Trends können dadurch frühzeitig erkannt werden.

Abschließend wird dem Leser ein Anwendungsfall namens „Restaurant Bot“ vorgestellt, der die Prozesse des Customer Services automatisiert und dadurch die Arbeit der Customer-Service-Mitarbeiter optimiert. Darüber hinaus werden die Vorteile der Einführung eines Chatbots in einem Restaurant, u.a. die Verbesserung der Unternehmen-Kunde-Interaktion sowie die Optimierung der Kundenwartezeit durch Vorbestellung und Erhöhung des Umsatzes, anhand eines Anwendungsbeispiels aufgezeigt.

# DataOps für cloudbasierte Artefakte und Anwendungen

Laura Leester

Prof. Dr. Harald Ritz

Dr. Bernd Heßbrügge

Technische Hochschule Mittel-  
hessen

Technische Hochschule Mittel-  
hessen

INFOMOTION GmbH

Fachbereich MNI  
Wiesenstr. 14  
35390 Gießen  
E-Mail: [laura.leester@mni.thm.de](mailto:laura.leester@mni.thm.de)

Fachbereich MNI  
Wiesenstr. 14  
35390 Gießen  
E-Mail: [harald.ritz@mni.thm.de](mailto:harald.ritz@mni.thm.de)

BU Strategy & Digital Solutions  
Westhafenplatz 1  
60327 Frankfurt  
E-Mail: [bernd.hessbruegge@infomotion.de](mailto:bernd.hessbruegge@infomotion.de)

## Kategorie

Bachelorarbeit

## Schlüsselwörter

DataOps, Cloud Computing

## Zusammenfassung

Mit wachsendem Interesse befassen sich deutsche Unternehmen mit der Thematik Datenspeicherung und dem damit verbundenen Datenmanagement. Hier wurde das Potential erkannt, durch datengetriebene Prozesse und Entscheidungen eine bessere Position im Wettbewerb zu erlangen und so nach außen und innen Vorteile für das Unternehmen erwirken zu können. Die Problemstellung umfasst die (unbekannten) Möglichkeiten, das Datenmanagement mit DataOps zu verbessern. Diese Thesis befasst sich speziell mit Unternehmen, die unter Einsatz von Cloud Computing Artefakte und Anwendungen nutzen oder produzieren. Die Thesis stellt auf Basis dieses Interesses im Verlauf die Untersuchung der Eignung der DataOps-Methode für cloudbasierte Artefakte und Anwendungen dar.

Um die Forschungsfrage zu beantworten werden zunächst Definitionen der verschiedenen Themengebiete aufgesetzt. Deren Eigenschaften werden extrahiert und in Bezug zueinander gesetzt, um Gemeinsamkeiten und Unterschiede der Ziele und Methoden der verschiedenen Themengebiete aufzuzeigen und eine Eignung und einen gemeinsamen Nutzungskontext abzuleiten. Die DataOps-Methode stellte sich als eine agile, datenbezogene Vorgehensweise für Rollen im Data Science-Bereich dar. Es fällt auf, dass diverse DataOps-Prinzipien darauf hinarbeiten, dass die Datenkompetenz eines Unternehmens erhöht wird, da DataOps Daten als wertschöpfenden Faktor betrachtet und diesen einen hohen Wert beimisst. Beiwerke zu DataOps sind die verwandten DevOps-, Agile- und Lean Manufacturing-Methoden. Davon übernimmt DataOps die Agilität mit kurzen Arbeitszyklen und dem Ziel, schnell Ergebnisse zu produzieren, um Kundenbedürfnisse abzudecken. Ähnlich wie DevOps auch Rollen an Mitarbeiter vergibt oder eine Struktur zwischen Teams aufstellt, so legt auch DataOps Rollen für Data Governance, Analysten und Ingenieure fest, um durch die übersichtliche Struktur die Kommunikation zwischen den Teams besser zu koordinieren. Anschließend erfolgt

die Betrachtung cloudbasierter Artefakte und Anwendungen. Artefakte treten in Unternehmen im Data-Science-Bereich in diversen Varianten auf und es ist wichtig, hier klar abzutrennen, dass von allen im Data-Science-Life-Cycle anfallenden Artefakten gesprochen wird, nicht nur jenen, die in der Modellierung und Entwicklung auftreten. Den Einsatz von Cloud Computing rechtfertigen dabei nicht alle Varianten der Artefakte, die als irrelevant betrachtet werden hier durch den Vergleich mit den Eigenschaften von Cloud-Services herausgestellt und für die weitere Betrachtung vernachlässigt. Cloudbasierte Anwendungen lassen sich dagegen relativ simpel auf bereits bekannte Cloud-Services zurückführen: das Software-as-a-Service-Angebot (SaaS) als direkt nutzbare Software sowie das Platform-as-a-Service-Angebot (PaaS) zur Softwareentwicklung in der Cloud. Die als cloud-relevant eingestuften Artefakte und Anwendungen werden mit den Eigenschaften und Anforderungen der DataOps-Methode verglichen. Hier zeigt sich, dass der Umgang mit cloudbasierten Artefakten durch den Einsatz von DataOps profitieren kann. Auch PaaS und SaaS wurden hierzu überprüft, wobei sich zeigt, dass SaaS losgelöst von der Einwirkung von DataOps ist, PaaS aber den Nutzern erlaubt, Einfluss auf Faktoren zu nehmen, die DataOps zu verbessern sucht. Der innovative Umgang mit Daten in auf PaaS bereitgestellte Testumgebungen oder beliebiges Einrichten von Schnittstellen werden nicht restringiert. Bei SaaS sind Anwender auf die Einstellungen und Konfigurationsmöglichkeiten angewiesen, die der Anbieter zur Verfügung stellt, ohne dass sie weiteres Customizing vornehmen können. Die Integration in vorhandene Systeme kann so deutlich komplexer ausfallen als bei PaaS-Systemen. So macht PaaS das Datenmanagement einfacher und individualisierbar. Die Data Governance kann besser eingehalten werden und unternehmensspezifische Ziele im Data&Analytics-Bereich werden mit PaaS leichter erreicht, als mit generischen SaaS-Angeboten. Auch zeigt sich, bei welchen cloudbasierten Artefakten und Anwendungen der Einsatz von DataOps einen Nutzen bringt und wie DataOps in diesem Fall eingesetzt werden kann. Es ist zu beachten, dass diese Thesis davon ausgeht, dass cloudbasierte Artefakte und Anwendungen eingesetzt werden, es erfolgt keine Evaluierung der generellen Nutzung von DataOps für klassische on-prem Data-Center.

# Einsatzmöglichkeiten von SAP Intelligent Robotic Process Automation anhand eines betrieblichen Anwendungsszenarios

Yulia Malysheva

Technische Hochschule  
Mittelhessen

Fachbereich Mathematik,  
Naturwissenschaften  
und Datenverarbeitung  
Wilhelm-Leuschner-Straße 13  
61169 Friedberg  
E-Mail:  
[yulia.malysheva@mnd.thm.de](mailto:yulia.malysheva@mnd.thm.de)

Prof. Dr. Harald Ritz

Technische Hochschule  
Mittelhessen

Fachbereich Mathematik,  
Naturwissenschaften und  
Informatik  
Wiesentraße 14  
35390 Gießen  
E-Mail: [harald.ritz@mni.thm.de](mailto:harald.ritz@mni.thm.de)

Peter Keiner

advades GmbH

Kaiserstraße 7  
35745 Herborn  
E-Mail: [p.keiner@advades.com](mailto:p.keiner@advades.com)

## Kategorie Bachelorarbeit

### Schlüsselwörter

Robotic Process Automation, SAP IRPA, Chatbots, SAP Conversational AI, Digitalisierung, Automatisierung

### Zusammenfassung

Heutzutage sind die Unternehmen aufgrund starker Konkurrenz und sich immer wandelnden Technologien verstärkt gefordert, sich mit technologischem Fortschritt auseinanderzusetzen und ihre Geschäftsprozesse zu optimieren, um Wettbewerbsvorteile zu schaffen. Ein Trend, der seit Jahren in Wirtschaft und Praxis diskutiert wird, ist die Automatisierung von Tätigkeiten in verschiedenen Unternehmensbereichen. Durch Effizienzdruck wird der Wunsch nach Automatisierung der Prozesse größer. Einige davon sind zeitintensive Routineaufgaben, die nach starren Mustern ablaufen. Ihre Automatisierung kann den Mitarbeitern ermöglichen, sich auf andere, wichtigere und kreative Aufgaben zu konzentrieren. Mit Robotic Process Automation (RPA) können Prozesse aufgenommen und automatisiert werden. Dabei steigt die Effizienz des gesamten Prozesses, da der Einsatz von vorprogrammierten Bots eine parallele Ausführung von mehreren Aufgaben ermöglicht und Fehler reduziert. Die Mitarbeiter können dann mehr in die wertschöpfenden Prozesse eingebunden werden. Neben etablierten, markt führenden RPA-Anbietern gewinnt auch die neue Software von SAP, Intelligent Robotic Process Automation (SAP IRPA) an Bedeutung, welche digitale Roboter in die firmeneigene SAP-Systemlandschaft integriert. Im Rahmen der Bachelorarbeit wurde die Software SAP IRPA untersucht und ihre Funktionsweise nach Kriterien wie Software-Bestandteile, Systemvoraussetzungen und Integrationsfähigkeit, Kosten, Support und Training sowie Kundenreferenzen untersucht. Weiterhin wurde die Software bewertet und mit der markt führenden Konkurrenz verglichen. Schließlich wurde der umgesetzte Anwendungsfall vorgestellt und bewer-

tet. Dieser zeigt die Verbindung des RPA-Bots mit einem Chatbot von der Plattform SAP Conversational AI.

Als Ergebnis wurde deutlich, dass RPA eine spannende Technologie mit viel Potential ist, mit der Unternehmen bestehende Prozesse mit repetitiven Aufgaben automatisieren können. RPA hat zahlreiche Vorteile, von der Kostenreduktion über Zeiteinsparung bis hin zu Qualitätssteigerung. Die genannten Vorteile können nur durch die richtige Automatisierungsstrategie erzielt werden. Zwar kann RPA in zahlreichen Anwendungsbereichen eingesetzt werden, aber nicht alle Prozesse eignen sich für die Automatisierung mittels RPA-Lösungen. Dafür müssen Prozesse bestimmten Auswahlkriterien entsprechen.

Standardisierte Prozesse mit wenigen Entscheidungen, einer hohen Frequenz und einem hohen Volumen, die idealerweise mit strukturierten Daten arbeiten, eignen sich am besten für die Automatisierung mit RPA. Erst bei der Erfüllung dieser Kriterien kann der Einsatz von RPA einen hohen ROI erzielen.

Mit SAP IRPA können Routineaufgaben automatisiert werden, sodass sich die Mitarbeiter auf höherwertige, wertschöpfende Tätigkeiten konzentrieren können. Wenn man das Werkzeug von SAP mit anderen Konkurrenzprodukten vergleicht, stellt man fest, dass SAP IRPA einen ähnlichen Funktionsumfang anbietet. Wie andere RPA-Werkzeuge hat die Lösung von SAP ein breites Spektrum an Automatisierungsmöglichkeiten, die von Web-, SAP- und Desktop-Anwendungen bis hin zu Microsoft-Office-Produkten reichen. Im Vergleich zu den drei Marktführern (UiPath, Automation Anywhere und Blue Prism) liegt der Schwerpunkt von SAP IRPA auf der Automatisierung von Geschäftsprozessen im SAP-Bereich. Eine einfache Integration in die bestehende SAP-Landschaft und ein flexibles Kostenmodell sind die Gründe, weshalb sich die Unternehmen für die RPA-Software von SAP entscheiden können. Aufgrund des noch jungen Alters der Lösung werden jedoch einzelne Funktionalitäten erst nach und nach eingeführt.

# Anforderungen an eine intermodale Mobilitätsplattform am Beispiel Berlin

Finn Honsberg (B.Sc.)

Prof. Dr. Birte Malzahn

Technische Universität München

HTW Berlin

Studiengang  
Wirtschaftsinformatik (M.Sc.)  
Arcisstraße 21  
80333 München  
finn.honsberg@tum.de

Studiengang  
Wirtschaftsinformatik  
Treskowallee 8  
10318 Berlin  
E-Mail: malzahn@htw-berlin.de

## Kategorie

Bachelorarbeit

## Schlüsselwörter

Intermodale Mobilität, urbane Mobilität, digitale Plattform, qualitativ-empirische Untersuchung

## Zusammenfassung

Verkehrssysteme in urbanen Räumen stoßen häufig an ihre Belastungsgrenze. Stark ausgelastet sind insbesondere die Straßen. Es werden deshalb Konzepte gesucht und entwickelt, um den individuellen motorisierten Verkehr hin zu anderen Mobilitätsformen zu lenken sowie insgesamt eine flexible und adaptive Auswahl von Verkehrsmitteln zu fördern.

In den letzten Jahren wurden Plattformen entwickelt, über die verschiedene Mobilitätsangebote wie Sharing-Angebote, Taxifahrten oder Angebote des öffentlichen Nahverkehrs offeriert werden. Eine *integrierte* Mobilitätsplattform bündelt dabei Mobilitätsangebote verschiedener Anbieter auf einer Plattform. Integrierte Plattformen verschaffen eine transparente Sicht auf verschiedene Mobilitätsangebote einer Stadt / eines Gebiets und können bisher unbekannte Mobilitätsalternativen zugänglich machen. Herkömmliche integrierte Plattformen bieten jedoch keine Kombination verschiedener Mobilitätsangebote im Rahmen einer Fahrtstrecke an.

Eine *intermodale* Mobilitätsplattform integriert ebenfalls die Mobilitätsangebote verschiedener Anbieter, und bietet darüber hinaus die Möglichkeit, die verschiedenen Mobilitätsangebote miteinander zu kombinieren, um eine ausgewählte Strecke zurückzulegen. So kann beispielsweise ein Bike-Sharing-Angebot mit dem öffentlichen Nahverkehr kombiniert werden. Die Kombination verschiedener Verkehrsmittel ermöglicht eine vorteilhaftere Routenplanung als eine herkömmliche Planung. In der Praxis sind solche Plattformen bisher jedoch nur in Ansätzen vorhanden.

Die vorliegende Bachelorarbeit eruiert Anforderungen, die Nutzer\_innen an eine solche intermodale Mobilitätsplattform stellen. Zu diesem Zweck werden zunächst theoretische Grundlagen zu den Themen

Mobilität, Verkehrsmodi und –mittel, On-Demand Mobilität und intermodale Mobilität gelegt. Da die Untersuchung am Beispiel der Stadt Berlin erfolgt, wird anschließend die Mobilitätssituation in Berlin beschrieben. Es zeigt sich u. a., dass bisher kaum Wege intermodal zurückgelegt werden.

Die empirische Datenerhebung erfolgt in Form von sechs qualitativen Tiefeninterviews, die mit Hilfe der Software MAXQDA 2020 ausgewertet werden. Die Ergebnisse zeigen, dass die Anforderungen an eine intermodale Mobilitätsplattform vielfältig sind:

- Routenplanung anhand von Präferenzen des Nutzers / der Nutzerin
- Anzeige der Auslastung von Verkehrsmitteln bzw. Verkehrswegen in Echtzeit, ggfs. mit Möglichkeit zum Wechsel einer bereits begonnenen Route
- Anzeige möglicher Parkflächen am Umstiegspunkt / Ziel
- Safe Mode: Auswahl sicherer Umstiegspunkte z. B. zur Nachtzeit
- Transparente Aufspaltung des Gesamtpreises in Komponenten
- Gemeinsame Abrechnung aller genutzten Services
- Möglichkeit von angebotsübergreifenden Abonnements / Paketpreisen

Ausschlaggebend für die Verwendung einer intermodalen Mobilitätsplattform ist grundsätzlich, dass sich für die Nutzer\_innen ein Mehrwert ergibt.

Die in der Arbeit erzielten Ergebnisse unterliegen folgenden Einschränkungen: Im Rahmen der qualitativen Tiefeninterviews wurden nur wenige Personen befragt. Diese wurden zudem gezielt nach vorher definierten Personas ausgewählt. Um verallgemeinerbare Ergebnisse zu erhalten, muss eine quantitative Studie durchgeführt werden. Des Weiteren beziehen sich die Daten nur auf den Berliner Raum. Schließlich wurden die Tiefeninterviews zur Zeit der COVID-19-Pandemie durchgeführt. Es ist denkbar, dass diese Situation die Ergebnisse beeinflusst hat. Die Arbeit kann als Basis für weiterführende Untersuchungen dienen.



# DEFINITION EINES VORGEHENSMODELLS ZUR NUTZENORIENTIERTEN ENTWICKLUNG DIGITALER PRODUKT-SERVICE-SYSTEME IN DER INVESTITIONSGÜTERINDUSTRIE AUF BASIS DER DESIGN SCIENCE RESEARCH METHODOLOGY

Philipp Mikat

Hochschule München  
Lothstraße 64  
80335 München

E-Mail: [philipp.mikat@gmail.com](mailto:philipp.mikat@gmail.com)

Jörg Puchan

Munich School of Engineering and Management  
Hochschule München  
Lothstraße 64  
80335 München

E-Mail: [puchan@hm.edu](mailto:puchan@hm.edu)

## KEYWORDS

Smart Services, Produkt-Service-System, Design Science Research, Digitalisierung, Industrie 4.0

## ABSTRACT

Der vorliegende Beitrag beschreibt ein Vorgehensmodell zur nutzenorientierten Entwicklung digitaler Produkt-Service-Systeme (PSS) in der Investitionsgüterindustrie. Methodische Grundlage hierfür ist die Design Science Research Methodology (DSRM).

## ZIELSTELLUNG

Während Entscheider in der Investitionsgüterindustrie bei Industrie 4.0 vorrangig Optimierungspotentiale für ihre Produktion und Prozesse betrachten,<sup>1</sup> bietet Industrie 4.0 auch die Möglichkeit, neue Geschäftsmodelle voranzutreiben. Ein Beispiel sind „as-a-Service“-Angebote<sup>2</sup>. Dazu gehören auch Smart Services<sup>3</sup>, die als vermarktbar, individuell auf die Bedürfnisse der Kunden ausgerichtete<sup>4</sup> Kombination aus Sach- und Dienstleistungen<sup>5</sup> sowie Software<sup>6</sup> häufig als Produkt-Service-Systeme bezeichnet werden.<sup>7</sup>

Es besteht ein Bedarf an einem Vorgehensmodell für Investitionsgüterhersteller, das es Anwendern ermöglicht, einen maximalen Nutzen beim Engineering von digitalen PSS zu erzielen. Dieser wird dann erreicht, wenn die Unternehmen im Entwicklungsprozess dazu befähigt werden, die sich bietenden Chancen zu nutzen und wesentliche Herausforderungen zu überwinden.

Um die Nutzenorientierung des Modells in der Praxis zu erhöhen, werden zudem relevante Erfolgsfaktoren definiert. Neben einer Literaturrecherche werden verschiedene Investitionsgüterhersteller mittels Sekundärrecher-

che analysiert. Dabei werden die folgenden Fragen beantwortet:

- Welche internen Stärken und externe Chancen erlauben es den Herstellern, Smart Services erfolgreich anbieten können?
- Welche internen Schwächen und externe Gefahren stellen die Unternehmen vor wesentliche Herausforderungen?

## VORGEHENSMODELL

Methodische Grundlage des Modells ist der von Peffers et al. geschaffene Prozess für die Forschung auf Basis der Design Science Research Methodology<sup>8</sup>, der um weitere Erkenntnisse aus dem Engineering von Smart Services und PSS ergänzt wird. Das Modell ist als Kreislauf konzipiert, wodurch es den Unternehmen ermöglicht werden soll, bisher gewonnene Erkenntnisse optimal zu nutzen. Neben einer Wissensbasis über Smart Services wird ebenso der Kunde als Wissensbasis eingeführt.<sup>9</sup>

Das Modell wird in Anlehnung an das Modell von Senderek et al. in drei Abschnitte unterteilt, wobei zunächst die strategische Zielsetzung erfolgt.<sup>10</sup> Hierbei wird ein spezifischer Anwendungsfall des Kunden konkret beschrieben und seine dahinterliegende Herausforderung konzeptionell konkretisiert.<sup>11</sup> Weiterhin erfolgt eine quantitative Beschreibung der Ziele des Produkts. Die definierten Erfolgsfaktoren werden zur Erhöhung des Gesamtnutzens als Ziele vorgegeben. Diese sind zudem ein Teil der Wissensbasis zu Smart Services.

<sup>1</sup>World Economic Forum 2015, S. 9.

<sup>2</sup>McKinsey & Company 2016, S. 22. Kagenmann et al., S. 14

<sup>3</sup>acatech – Deutsche Akademie der Technikwissenschaften o. J.

<sup>4</sup>Buriánek 2009, S. 38.

<sup>5</sup>Goedkoop 1999, S. 18.,

<sup>6</sup>Meier und Uhlmann 2012, S. 6.

<sup>7</sup>Abramovici 2018, S. 4, Beverungen et al. 2017, S. 784–785.

<sup>8</sup>Peffers et al. 2007, S. 52–56.

<sup>9</sup>Niemöller et al. 2014, S. 188, DIN SPEC 33453:2019-09, S. 14.

<sup>10</sup>Senderek et al. 2019, S. 6–12.

<sup>11</sup>Grohmann et al. 2017, S. 26.

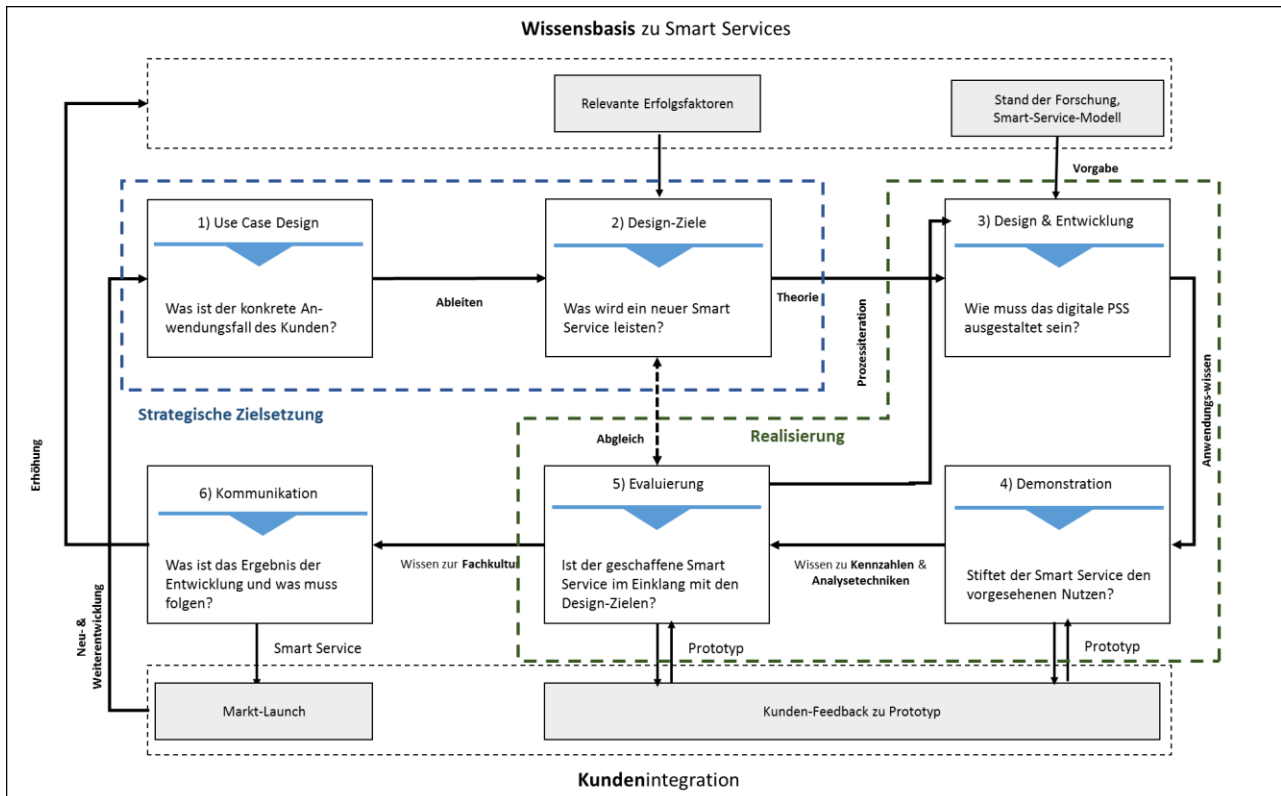


Abbildung 1<sup>12</sup>: Vorgehensmodell zur nutzenorientierten Entwicklung von digitalen PSS.

Im zweiten Abschnitt erfolgt die Realisierung des Smart Services.<sup>13</sup> Konkret werden zunächst die Funktionalitäten und die Architektur des digitalen PSS bestimmt sowie die Demonstration und die formale Evaluierung durchgeführt. Im Zuge der Evaluierung werden die Funktionalitäten des tatsächlich vorliegenden Smart-Service-Prototyps mit den definierten Design-Zielen verglichen.<sup>14</sup> Das Modell gibt vor, dass dieser Abgleich durch den Kunden und unternehmensinterne Experten erfolgen muss. Eine positive Rückmeldung durch die Evaluierenden ist zwingend erforderlich, um zur nächsten Aktivität voranzuschreiten.<sup>15</sup> Bei Nicht-Erreichung kann das Unternehmen beliebig häufig eine Iteration beginnend ab der Realisierungsphase durchführen.<sup>16</sup>

In dritten Abschnitt erfolgt die Kommunikation des Artefakts. Zum einen erfolgt sie unternehmensintern um die notwendige Entwicklung der Organisation vollziehen und die Wissensbasis zu Smart Services weiter ausbauen zu können.<sup>17</sup> Zum anderen wird das PSS an externe Adressaten kommuniziert: Das Produkt wird im Markt eingeführt.<sup>18</sup> Neben der Erfüllung vertrieblicher Ziele dient dies auch dazu, das Wissen des Kunden über angebotene

Smart Services zu erhöhen,<sup>19</sup> was wiederum den Ausgangspunkt für weitere Engineering-Projekte darstellt.

### BEWERTUNG DES MODELLS

Das Modell sowie die Erfolgsfaktoren wurden in erster Iteration im Rahmen einer Evaluierung durch einen Experten für PSS aus der Praxis positiv bewertet. Die hohe Relevanz der Erfolgsfaktoren sowie deren Unterstützung hin zu einer gesamtheitlichen Entwicklung des angestrebten PSS werden hervorgehoben. Gleichzeitig wird angeregt, dass in der Realisierungsphase Lösungsalternativen und ein Entscheidungsmodell generiert werden sollten.

### FAZIT

Es zeigt sich, dass in der Theorie mit Hilfe der Design Science Research Methodology ein komplexes, markt-orientiertes Produkt agil entwickelt werden kann. Ferner zeigt die Evaluierung durch einen Experten für PSS, dass die Kernanforderung hinsichtlich der Erreichung eines maximalen Nutzens bei der Entwicklung von digitalen PSS durch dieses Modell gegeben ist.

<sup>12</sup> eigene Darstellung in Anlehnung an: Peffers et al. (2007), Exner et al. (2014, S. 70–71), Senderek et al. (2019, S. 9-12), DIN SPEC 33453:2019-09:2019 (S. 14-15), Niemöller et al. 2014, S. 188, Stark et al. (2009, S. 4–6), Hevner et al. (2004, S. 80-90), Grohmann, Jungmann und Wambacher (2017, S. 25-32), Bonny et al. (2019, S. 6), Hevner (2007, S. 4), acatech – Deutsche Akademie der Technikwissenschaften (o. J.).

<sup>13</sup> Senderek et al. 2019, S. 9–10.

<sup>14</sup> Peffers et al. 2007, S. 56.

<sup>15</sup> Senderek et al. 2019, S. 10, DIN SPEC 33453:2019-09, S. 14.

<sup>16</sup> Peffers et al. 2007, S. 56, DIN SPEC 33453:2019-09, S. 14.

<sup>17</sup> Hevner et al. 2004, S. 90, Hevner 2007, S. 4.

<sup>18</sup> DIN SPEC 33453:2019-09, S. 15, Senderek et al. 2019, S. 10–12.

<sup>19</sup> Hevner 2007, S. 4, Hevner et al. 2004, S. 90.

Die ermittelten Erfolgsfaktoren weisen durch Unternehmensanalyse einen hohen Praxisbezug auf. Um weitere für den Erfolg von Smart Services entscheidende Faktoren zu ermitteln, ist die Durchführung von Entwicklungsprojekten bei Investitionsgüterherstellern zu sehen. Auf Basis der gewonnenen Erkenntnisse sollen weitere Optimierungspotentiale und Handlungsfelder identifiziert und bestehende Optimierungspotentiale ausgeschöpft werden.

Relevante Erfolgsfaktoren	
1)	<b>Strategie-Alignment</b> Das entwickelte digitale Produkt-Service-System muss im Einklang mit einer Digitalisierungsstrategie des Unternehmens sein.
2)	<b>Kernkompetenz</b> Für die Entwicklung des Smart Services muss das gewonnene digitale Wissen mit der Kernkompetenz des Unternehmens kombiniert werden.
3)	<b>Kundennutzen</b> Für den Kunden muss ein Nutzen generiert werden, der an seinen Erwartungen orientiert ist.
4)	<b>Anbiertennutzen</b> Für das herstellende Unternehmen muss der vorliegende Smart Service einen Mehrwert erbringen, beispielsweise in Form von einer Ausschöpfung von Potentialen zur Prozessoptimierung.
5)	<b>Sicherheitskonzept</b> Es muss ein funktionierendes Sicherheitskonzept implementiert und kontinuierlich weiterentwickelt werden.
6)	<b>Datenhoheit</b> Der Hersteller muss die Nutzungsverfügung über die generierten, nicht-personenbezogenen Daten erlangen.

Abbildung 2: Die relevanten Erfolgsfaktoren (eigene Darstellung).

## LITERATUR

Abramovici, Michael (Hg.) (2018): Engineering smarter Produkte und Services Plattform Industrie 4.0 STUDIE. acatech – Deutsche Akademie der Technikwissenschaften. München.

acatech – Deutsche Akademie der Technikwissenschaften (Hg.) (o. J.): Smart Service Welt – Internetbasierte Dienste für die Wirtschaft. Online verfügbar unter <https://www.acatech.de/projekt/smart-service-welt/>, zuletzt geprüft am 14.03.2020.

Beverungen, Daniel; Matzner, Martin; Janiesch, Christian (2017): Information systems for smart services. In: *Inf Syst E-Bus Manage* 15 (4), S. 781–787. DOI: 10.1007/s10257-017-0365-8.

Buriánek, Ferdinand (2009): Vertragsgestaltung bei hybriden Leistungsangeboten. Eine ökonomische Betrachtung. Dissertation Technische Universität München, 2009. 1. Aufl. Wiesbaden: Gabler Verlag (Markt- und Unternehmensentwicklung Markets and Organisations). Online verfügbar unter <http://gbv.ebib.com/patron/FullRecord.aspx?p=748058>.

DIN SPEC 33453:2019-09, 2019: Entwicklung digitaler Dienstleistungssysteme.

Goedkoop, Mark Jacob (1999): Product Service systems, Ecological and Economic Basics.

Grohmann, Alexander; Jungmann, Michael; Wambacher, Roman (2017): Smart Products und Smart Services

entwickeln. Herausforderungen & Erfolgsfaktoren. In: Arndt Borgmeier, Alexander Grohmann und Stefan F. Gross (Hg.): Smart services und Internet der Dinge. Geschäftsmodelle, Umsetzung und Best Practices : Industrie 4.0, Internet of Things (IoT), Machine-to-Machine, Big Data, Augmented Reality Technologie. München: Hanser, S. 23–38. Online verfügbar unter <http://dx.doi.org/10.3139/9783446452701>.

Hevner, Alan R. (2007): A Three Cycle View of Design Science Research. In: *Scandinavian Journal of Information Systems* 19 (1).

Hevner, Alan R.; March, Salvatore T.; Park, Jinsoo; Ram, Sudha (2004): Design Science in Information Systems Research. In: *MIS Quarterly* 28 (1), S. 75–104.

Kagermann, Henning; Riemensperger, Frank; Hoke, Dirk; Schuh, Günther; Scheer, August-Wilhelm; Spath, Dieter et al. (Hg.): Smart Service Welt. Umsetzungsempfehlungen für das Zukunftsprojekt internetbasierte Dienste für die Wirtschaft. Abschlussbericht. Langversion. Berlin.

McKinsey & Company (Hg.) (2016): Industry 4.0 after the initial hype. Where manufacturers are finding value and how they can best capture it. Unter Mitarbeit von Dominik Wee, Matthias Breunig und Valerie von der Tann. McKinsey & Company. Online verfügbar unter [https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/getting%20the%20most%20out%20of%20industry%204%200/mckinsey\\_industry\\_40\\_2016.ashx](https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/getting%20the%20most%20out%20of%20industry%204%200/mckinsey_industry_40_2016.ashx), zuletzt geprüft am 16.04.2020.

Meier, Horst; Uhlmann, Eckart (Hg.) (2012): Integrierte Industrielle Sach- und Dienstleistungen. Vermarktung, Entwicklung und Erbringung hybrider Leistungsbündel. Berlin, Heidelberg: Springer Berlin Heidelberg.

Niemöller, Christina; Özcan, Deniz; Metzger, Dirk; Thomas, Oliver (2014): Towards a Design Science-Driven Product-Service System Engineering Methodology. In: Monica Chiarini Tremblay, Debra VanderMeer, Marcus Rothenberger, Ashish Gupta und Victoria Yoon (Hg.): Advancing the Impact of Design Science: Moving from Theory to Practice. 9th International Conference, DESRIST 2014, Miami, FL, USA, May 22-24, 2014. Proceedings. Cham: Springer International Publishing; Imprint; Springer (Lecture Notes in Computer Science, 8463), S. 180–193.

Peppers, Ken; Tuunanen, Tuure; Rothenberger, Marcus A.; Chatterjee, Samir (2007): A Design Science Research Methodology for Information Systems Research. In: *Journal of Management Information Systems* 24 (3), S. 45–77. DOI: 10.2753/MIS0742-122240302.

Senderek, Roman; Ragab, Sherif; Stratmann, Lukas; Krechting, Denis (2019): Smart-Service-Engineering. Eine agile Herangehensweise zur Entwicklung datenbasierter Services. In: Volker Stich, Jan Hendrik Schumann, Daniel Beverungen, Gerhard Gudergan und Philipp Jussen (Hg.): Digitale Dienstleistungsinnovationen. Berlin, Heidelberg: Springer Berlin Heidelberg.

World Economic Forum (2015): Industrial Internet of Things. Unleashing the Potential of Connected Products and Services. Hg. v. World Economic Forum. World Economic Forum. Genf (020315). Online verfügbar unter [http://www3.weforum.org/docs/WEFUSA\\_IndustrialInternet\\_Report2015.pdf](http://www3.weforum.org/docs/WEFUSA_IndustrialInternet_Report2015.pdf), zuletzt geprüft am 04.05.2020.

# End-to-end Data-Warehouse-Szenario von Social-Media-Daten mithilfe moderner SAP-Technologien

Frederic Wall

Hochschule Pforzheim  
Tiefenbronner Straße 65  
75175 Pforzheim  
frederic.a.w@web.de

Oliver Meier

SAP SE  
Dietmar-Hopp-Allee 16  
69190 Walldorf

Frank Morelli

Hochschule Pforzheim  
Tiefenbronner Straße 65  
75175 Pforzheim  
frank.morelli@hs-pforzheim.de

## Schlüsselwörter

Sentimentanalyse, Data Warehousing, Social-Media-Daten, Twitter, Qualtrics

## Problemstellung und Zielsetzung

Weltweit versuchen Unternehmen Social-Media-Daten auszuwerten um wertvolle Informationen zu erhalten. Sie nutzen diese, um bspw. Kosten durch optimierte Werbekampagnen einzusparen, neue Erkenntnisse über die Kunden zu erhalten oder um eine optimierte Informationslage für zukünftige Unternehmensentscheidungen zu schaffen. Ein Problem bei der Verarbeitung und Auswertung von Daten aus sozialen Netzwerken besteht darin, dass ein großer Teil der Daten unstrukturiert und in heterogenen Formaten vorliegt, was eine automatisierte Verarbeitung erschwert. Im Rahmen der Masterarbeit wird ein End-to-End Data Warehouse Szenario auf Basis mehrerer SAP Tools (HANA, BW/4HANA, SAP Analytics Cloud) entwickelt.

Als Use Case fungiert eine Sentimentanalyse. Diese ermöglicht die automatisierte Auswertung von Texten aus sozialen Netzwerken, bspw. Twitter. Mithilfe diverser Softwarelösungen kann die Extraktion, Speicherung und Auswertung von Tweets automatisiert erfolgen. Die Ergebnisse der Sentimentanalyse können anschließend quantifiziert und im Kontext interner Unternehmensdaten sowie externer Umfrageergebnisse von Qualtrics visualisiert werden.

## Methodisches Vorgehen

Um ein anwendungsrelevantes Ergebnis zu erzielen erfolgt im Rahmen der Thesis die Konzeption eines Use Case mit betriebswirtschaftlicher und technischer Ausgangssituation. Anhand des Anwendungsfalls werden zwei Systemarchitekturalternativen aus einer vorgegebenen Anzahl an Softwaretools abgeleitet, welche die für das Erreichen der Zielsetzung geeignet sind. Um die vergleichsweise optimale Systemarchitektur für den gewählten Use Case zu bestimmen, erfolgt eine Evaluation der beiden Szenarien anhand von fünf Kriterien.

## Sentimentanalyse

Bei einer Sentimentanalyse handelt es sich um eine Reihe von Methoden, Techniken und Werkzeugen zur Erkennung und Extraktion subjektiver Informationen, bspw. Meinungen und Haltungen, aus natürlicher Sprache. Sentimentanalysen können in drei Ebenen erfolgen: Dokumentenebene, Satzebene und Aspektenebene. Während die Dokumentenebene klassifiziert, ob ein Dokument als Ganzes ein positives oder negatives Sentiment zum Ausdruck bringt, werden bei der Satzebene einzelne Sätze ausgewertet. Auf Aspektenebene erfolgt die Auswertung, welchen Aspekt die Ersteller des Dokuments negativ bzw. positiv bewerten.

In der Wissenschaft haben sich diverse technische Ansätze etabliert, um die beschriebene Auswertung natürlicher Sprache durchzuführen. Grundsätzlich lassen sich diese Vorgehensweisen in zwei Kategorien einteilen: technische Ansätze basierend auf Machine Learning vs. lexikalische Herangehensweisen. Die Technik der Sentimentanalyse auf SAP HANA ist dem wörterbuchbasierten Ansatz zuzuordnen. Hierbei werden Texte nach Meinungswörtern durchsucht, um ein hinterlegtes Lexikon nach Synonymen und Antonymen abzugleichen. Den Einträgen im Wörterbuch wird jeweils ein positives bzw. negatives Sentiment zugeordnet. Findet das System beim Durchsuchen des Texts zugehörige Einträge aus dem Lexikon, kann anhand des hinterlegten Sentiments auf die Stimmung des Texts geschlossen werden.

## Architekturvergleich

Für die Extraktion der Tweets, die Durchführung der Sentimentanalyse und die Visualisierung der Daten erweist sich die Kombination unterschiedlicher Softwarewerkzeuge als notwendig. Ausgearbeitet werden zwei Systemarchitekturen, wobei Variante A im Gegensatz zu Variante B kein Data Warehouse für die zentrale Datenkonsolidierung verwendet.

Bei Architektur A erfolgt der Import der Daten von Twitter mithilfe des Data Provisioning Agents in SAP HANA, wo die Datenaufbereitung und die Sentimentanalyse durchgeführt wird. Die Bereitstellung der Ergebnisse der Sentimentanalyse gemeinsam mit internen Finanzdaten wird in SAP Analytics Cloud (SAC) umgesetzt. Auf dieser Ebene werden ebenfalls

Umfrageergebnisse von Qualtrics mit den übrigen Daten harmonisiert und visualisiert. Abbildung 1 stellt den schematischen Aufbau der Architekturvariante A dar.

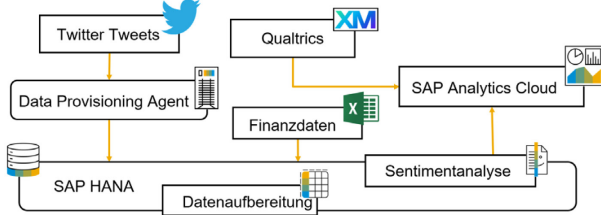


Abbildung 1: Systemarchitektur A

Bei Architektur B wird neben den Tools von Architektur A zusätzlich ein Data Warehouse eingebunden. Dieses ermöglicht die zentrale Harmonisierung und Konsolidierung aller Datenquellen und die anschließende Bereitstellung.

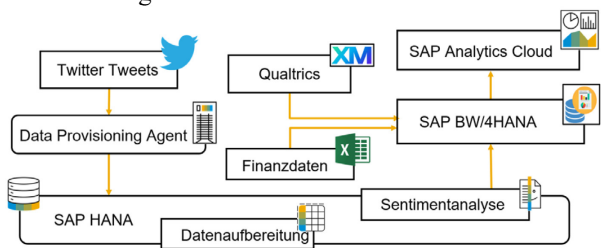


Abbildung 2: Systemarchitektur B

Um eine Entscheidung bzgl. der Architekturalternativen zu treffen, werden beide Ansätze anhand von fünf Bewertungskriterien qualitativ verglichen.

**Einfachheit** (bewertet die Vermeidung von Komplexität bei Implementierung, Modifizierung und Wartung der Architekturmodelle)

**Datenintegration** (evaluiert die Möglichkeit der Integration diverser Datenquellen und deren Modellierung)

**Flexibilität** (bewertet die Möglichkeit der Anpassung und Erweiterung der Architektur bzw. des Use Case)

**Stabilität** (beurteilt die Robustheit der einzelnen Architekturelemente sowie deren Integration)

**Compliance:** (evaluiert die Möglichkeit der Einhaltung von Gesetzen sowie interner und externer Richtlinien)

Die Evaluation erfolgt mithilfe einer 5-stufigen Harvey-Ball-Skala (○: nicht vorhanden, ◐: schwach ausgeprägt, ◑: mäßig ausgeprägt, ◒: deutlich ausgeprägt, ◓: erheblich ausgeprägt). In Tabelle 1 ist das Ergebnis des Architekturvergleichs dargestellt.

Kriterien	Architektur A	Architektur B
<b>Einfachheit</b>	◑	◐
<b>Datenintegration</b>	◐	◓
<b>Flexibilität</b>	◐	◒
<b>Stabilität</b>	◒	◒
<b>Compliance</b>	◐	◓

Tabelle 1: Architekturvergleich

Die Evaluation beider Systemarchitekturen weist auf die Vorteilhaftigkeit von Variante B hin: Zwar weist diese Alternative eine geringfügig höhere Komplexität aus, sie

ist Variante A jedoch in den anderen Bewertungskriterien überlegen.

### Kritische Würdigung

Während die Sentimentanalyse mithilfe des wörterbuchbasierten Ansatzes viele Sentimente richtig klassifizieren kann, stößt die Technologie teilweise auch an ihre Grenzen. Wie die meisten Technologien für Sentimentanalysen enthält SAP HANA keine Möglichkeit, ironische oder sarkastische Aussagen von Twitter-Nutzern zu erkennen, entsprechende Tweets werden daher in der Regel falsch zugeordnet. Darüber hinaus können Twitter-Bots, d.h. automatisierte Accounts, ebenfalls Einfluss auf die Ergebnisse der Sentimentanalyse nehmen.

Trotz vereinzelt inkorrekt klassifizierter Tweets lässt sich nach umfangreichen Stichproben die Annahme treffen, dass die Analyse meist einen korrekten Eindruck der Stimmung auf Twitter zu den entsprechenden Themen vermittelt.

### Fazit

Mithilfe der vorgestellten Architektur lassen sich Daten von Twitter automatisiert extrahieren, auswerten und eine Sentimentanalyse durchführen. Des Weiteren können Umfragewerte von Qualtrics direkt in ein Data Warehouse importiert und gemeinsam mit weiteren Unternehmensdaten konsolidiert werden. Mithilfe von SAC ist es beispielweise möglich, die Ergebnisse zu visualisieren und Endnutzern in Form von dynamischen Dashboards zur Verfügung zu stellen. Die erarbeitete Lösung stellt einen ersten Prototypen für die technische Umsetzung der Zielsetzung dar und kann weiter optimiert und erweitert werden. Die Umsetzung der vorgestellten Lösung vermag es, Unternehmen in die Lage zu versetzen, Daten sozialer Netzwerke automatisiert zu extrahieren, zu verarbeiten und zu interpretieren, um die Meinungen der Nutzer dieser Netzwerke auszuwerten und darauf zu reagieren.

# Intelligente Optimierung der langfristigen Liquiditätsplanung

Robin Schmid

Hochschule Pforzheim  
Tiefenbronner Straße 65  
75175 Pforzheim  
schmidro@hs-pforzheim.de

Frank Morelli

Hochschule Pforzheim  
Tiefenbronner Straße 65  
75175 Pforzheim  
frank.morelli@hs-pforzheim.de

## Schlüsselwörter

Liquiditätsplanung, Werttreiberplanung, Künstliche Intelligenz, Machine Learning, Predictive Analytics

## Problemstellung und Zielsetzung

Während Unternehmen eine längere Zeit ohne die Erzielung von Gewinnen überleben können, dauert es hingegen in Extremfällen nur wenige Tage, um zahlungsunfähig zu werden. Nur mit einem modernen Liquiditätsmanagement kann eine angemessene Absicherung von Liquiditätsschwankungen gewährleistet und Zeiten knapper Liquidität zuverlässig und kosteneffizient bewältigt werden. Der Nachteil langfristiger Planungen ist, dass man Prognosen und Vorhersagen nur mit hoher Unsicherheit tätigen kann. Der hohe manuelle Aufwand der Datenerhebung und -validierung passt meist nicht zum Verhältnis von Aufwand und Prognosegüte. Viele Unternehmen erstellen eine Planung noch immer mit einem (Punkt-)Wert bezogen auf die zu planenden Positionen. Das zentrale Problem gestaltet sich dabei wie folgt: Während die Vergangenheit sich mit einer Zahl oder einem (Punkt-)Wert darstellen lässt, ist dies für die Zukunft nicht möglich. Aufgrund der Ungewissheit der Zukunft, ist diese lediglich mit Bandbreiten adäquat abzubilden. Planungen und Prognosen werden heutzutage -unter der „VUCA (Volatility, Uncertainty, Complexity, Ambiguity)-World“ zunehmend komplexer. Die Auswirkungen auf die Liquiditätsplanung sind weitreichender als vermutet. Darüber hinaus herrscht oftmals fehlende Transparenz in Unternehmen über die wesentlichen Cashflow-Treiber, was die Liquiditätsplanung zusätzlich erschwert. In der heutigen Zeit wachsen die Datenmengen aufgrund der voranschreitenden Digitalisierung explosionsartig an. Mit dem Ziel, Big Data sowohl verwalten als auch für geschäftliche Zwecke nutzbar machen zu können, führt diese Entwicklung zu einer Trendwende hin zum Einsatz von Machine Learning (ML) und Predictive Analytics in Unternehmen. Mit Hilfe moderner statistischer Algorithmen sowie deren Einsatz in der Liquiditätsplanung lassen sich Muster, Trends und Strukturen genauer erkennen. Die Digitale Transformation und die Weiterentwicklung Künstlicher Intelligenz eröffnen völlig neuen Möglichkeiten und Herausforderungen im Unternehmens- und Wettbewerbsumfeld. Dies führt auch im Bereich des Cash-

und Liquiditätsmanagements zu neuen Gestaltungspotenzialen. Mit dem Einsatz Künstlicher Intelligenz, ist es zunehmend möglich, die vergangenen Treiber der Liquidität besser zu verstehen und darauf aufbauend die Cash- und Liquiditätsplanung neu zu organisieren. Diese Arbeit verfolgt das Ziel, die Frage zu beantworten, wie Künstliche Intelligenz (KI) dazu beitragen kann den langfristigen Planungsprozess im Liquiditätsmanagement zu verbessern.

## Optimierungspotenziale durch den Einsatz von KI

Traditionelle Planung und Prognosen haben sich aufgrund stetig steigender Datenmengen zu einem schwerfälligen und zeitintensiven Prozess entwickelt. Trotz vorhandener leistungsstarker IT wird oft auf die eigene Intuition und das eigene Urteilsvermögen zurückgegriffen. Dies gibt unbewussten Vorurteilen und bewussten Manipulationen Raum. Mit dem Einsatz von datengetriebenen, prädiktiven Algorithmen bietet sich Unternehmen die Möglichkeit das beschriebene Dilemma zu lösen. Ermöglicht wird dies durch neue Technologien wie In-Memory-Computing, Analytics Plattformen und den Einsatz von ML als Werkzeug der KI. Der zugehörige Wertbeitrag in der Unternehmensplanung ist sowohl qualitativer als auch quantitativer Natur. Als resultierende Wertbeiträge ergeben sich hieraus u.a.:

**1) Steigerung der Effizienz und Genauigkeit in der Planung:** Forecasts und operative Planungen können aufgrund von statistischen Datenanalysen von Advanced Analytics-Methoden schneller und mit einer höheren Genauigkeit durchgeführt werden.

**2) Rationalere Entscheidungen:** Algorithmen treffen Entscheidungen ausschließlich auf Basis von Fakten und sind bei ihren Prognosen konsistent.

**3) Quantifizierung und Analyse von Wirkungszusammenhängen:** Oftmals weisen Zusammenhänge, die aus den Erfahrungen von Mitarbeitern entstanden sind, ein hohes Abstraktionsniveau auf. Diese sind nur selten mit relevanten Daten untermauert. Auf Basis wirtschaftlicher Daten, unterstützen Advanced Analytics Methoden dabei, die Wirkungszusammenhänge tiefer und präziser zu ermitteln und bisher noch unbekannte Wirkungszusammenhänge aufzudecken.

## Intelligente Optimierung der langfristigen Liquiditätsplanung

Im Hinblick auf die Mängel und Herausforderungen traditioneller Planungsansätze sowie technologischer Fortschritte, erfordert es eine neue, moderne Art der Liquiditätsplanung. Die Entwicklung einer digitalen und intelligenten Liquiditätsplanung kann dabei in **drei Schritten** erfolgen. Die *erste Evolutionsstufe* ist die Umstellung der traditionellen Liquiditätsplanung auf eine treiber- und datenbasierte Planung mit internen Daten (Punktwert). Im Rahmen der treiberbasierten Planung werden relevante Prozesse mit Hilfe eines Treibermodells abgebildet. Abbildung 1 zeigt einen beispielhaften Aufbau eines Werttreibermodells, das die Veränderung der liquiden Mittel als finanzielle Ergebnisgröße betrachtet:

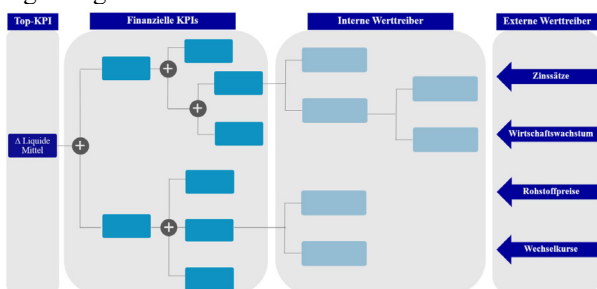


Abbildung 1: Aufbau eines Werttreiberbaums.

Ein Treibermodell lässt sich als Kennzahlensystem interpretieren, welches die betrachtete finanzielle Ergebnisgröße (z. B. EBIT, Liquide Mittel oder Free Cashflow) und die sie maßgeblich beeinflussenden Werttreiber miteinander in Verbindung setzt. Wirkungszusammenhänge zwischen finanziellen und operativen Größen werden damit in den Fokus gerückt. Als Resultat einer Werttreiberplanung werden Ergebnisgrößen nicht mehr direkt, sondern über den Einfluss maßgeblicher Werttreiber geplant. Dabei spaltet man die betrachtete Ergebnisgröße systematisch in ihre Bestandteile auf und verknüpft diese mit ihren jeweiligen Einflussfaktoren. Dies ermöglicht eine automatische Ableitung finanzieller Auswirkungen und eine stärkere Fokussierung auf die wesentlichen Werttreiber. In einem *zweiten Schritt* werden die wesentlichen Werttreiber durch den Einsatz von Advanced Analytics, auf Basis interner und externer Daten, ermittelt und prognostiziert (erste Bandbreiten und Korrelationen). Predictive Analytics in Verbindung mit Werttreibermodellen repräsentieren die Kernelemente einer modernen effektiveren und effizienteren Steuerung. Mit dem Einsatz von Predictive Analytics im Rahmen der Liquiditätsplanung werden qualitativ-theoretische Ursache-Wirkungsbeziehungen schrittweise durch quantitativ-statistische und datenbasierte Zusammenhänge ersetzt. Durch eine kontinuierliche Validierung der Zusammenhänge verbessert sich nach und nach die Abbildung der Realität in Treibermodellen. Im Rahmen der Liquiditätsplanung bedeutet Predictive Analytics das Prognostizieren von liquiditätsrelevanten

Treibern, mittels analytischer Hochrechnung historischer Daten. Durch das Einbeziehen von internen und externen Daten wird eine datengetriebene Vorhersage der Liquidität ermöglicht. Ein ML-Modell berücksichtigt und lernt dabei verschiedene Faktoren und deren Auswirkung in der Vergangenheit auf den zu prognostizierenden Cashflow. Durch Analysemethoden wie Klassifikations- und Regressionsanalysen, lässt sich neben dem Einfluss der Werttreiber auf die zu prognostizierende Größe auch die Stärke des Einflusses ermitteln. Schritt für Schritt können Werttreiber und Kennzahlen, die manuell mit hohem Aufwand, oder nur mit hoher Fehlerquote dargestellt werden können, mit Methoden der Predictive Analytics prognostiziert werden.

Die Planung in Szenarien durch Simulationen und die Integration von Wahrscheinlichkeiten, Chancen und Risiken ist der *dritte Schritt* auf dem Weg zur Optimierung der langfristigen Liquiditätsplanung (Bandbreiten, Korrelationen, automatische Analysen). Aufbauend auf einem besseren Verständnis für liquiditätsrelevante Treiber, lässt sich eine szenarioorientierte Planung idealerweise in Verbindung mit einem werttreiberbasierten Planungsmodell erstellen, das die wichtigsten Handlungsmöglichkeiten im Liquiditätsmanagement aufzeigt. Der größte Mehrwert einer Werttreiberplanung entsteht dann, wenn man Advanced Analytics Technologien zur Simulation verschiedener Werttreiber einsetzt. Kausale Zusammenhänge zwischen Zielgröße und den wichtigsten Stellhebeln eines Geschäftsmodells werden dadurch erkennbar. Auf diese Weise können auf Knopfdruck verschiedene künftige Entwicklungen (Szenarien) simuliert werden.

## Fazit

Mit dem Einsatz von Predictive Analytics in der Liquiditätsplanung lassen sich eine Vielzahl von Optimierungspotenzialen erzielen. Mit einer auf Werttreibern aufgebaute Liquiditätsplanung in Kombination mit dem Einsatz von ML-Techniken, können wesentliche Liquiditätstreiber nicht nur bestimmt, sondern auch prognostiziert werden. Zusätzlich ermöglicht dieser Planungsansatz den Einsatz von Simulationen, um automatisiert die Auswirkungen verschiedener Szenarien auf die Liquidität zu ermitteln und entsprechende Maßnahmen ableiten zu können. Eine Weiterentwicklung des Einsatzes von Predictive Analytics wäre z. B. die zusätzliche Integration externer Werttreiber, wie makroökonomische Faktoren, Preise für Rohstoffe, oder Informationen zu Wettbewerbern. ML-Modelle eignen sich besonders dazu, große Mengen an Daten gleichzeitig zu analysieren und für automatisierte Prognosen zu nutzen. So können die Planungs- und Prognoseprozesse, die bisher weitestgehend auf internen Unternehmensdaten basieren, durch externe Parameter ergänzt werden, um die Prognosequalität zu steigern und Ursache-Wirkungs-Beziehungen exakter zu identifizieren.



# Aufzeichnung und Wiedergabe einer 3D-Simulation bei der virtuellen Inbetriebnahme

Felix Richter (BSc.)

F.EE GmbH  
Abteilung Simulation und  
virtuelle Inbetriebnahme

In der Seugn 10  
92431 Neunburg v.W.  
E-Mail: [felix.richter@fee.de](mailto:felix.richter@fee.de)

Dipl.-Ing. (FH)  
Peter Meier

F.EE GmbH  
Abteilung Simulation und  
virtuelle Inbetriebnahme

In der Seugn 10  
92431 Neunburg v.W.  
E-Mail: [peter.meier@fee.de](mailto:peter.meier@fee.de)

Professor Dr.  
Frank Herrmann

OTH Regensburg  
Innovationszentrum für  
Produktionslogistik und  
Fabrikplanung

Galgenbergstraße 32  
93053 Regensburg  
E-Mail: [frank.herrmann@oth-regensburg.de](mailto:frank.herrmann@oth-regensburg.de)

## Kategorie

Bachelorarbeit

## Schlüsselwörter

Virtuelle Inbetriebnahme, digitaler Zwilling, Simulation, Aufnahme, paralleles Rechnen, Datenkompression, 3D, C#, .Net

## Zusammenfassung

Immer komplexere Industrieanlagen erfordern eine immer umfangreichere virtuelle Inbetriebnahme (VIBN). Parallel zur Softwareentwicklung der Steuerungsprogramme wird diese anhand einer 3D-Simulation der Anlage durchgeführt.

Während der VIBN muss der digitale Zwilling in bestimmten Situationen aufgezeichnet werden. Bei der F.EE GmbH wird die VIBN mit der firmeneigenen 3D-Simulationssoftware fe.screen-sim durchgeführt.

Zur Aufzeichnung von Simulationsabläufen wird bisher ein Videorecorder verwendet, welcher zwischen die Grafikkarte des Rechners und den Monitor gesteckt wird. Da bei dieser Methode der Blickwinkel bereits bei der Aufzeichnung festgelegt werden muss, ist die Wiedergabe auf diesen beschränkt. Dadurch werden beispielsweise Fehlerquellen außerhalb des Blickwinkels bei der VIBN nicht aufgezeichnet. Die Beschränkung auf einen Blickwinkel macht das genaue Betrachten von komplexen Simulationsprozessen unmöglich.

In dieser Bachelorarbeit wurde zur Lösung der oben genannten Problematik der in fe.screen-sim integrierte Recorder SimReplay entwickelt, welcher direkt 3D-Daten aus der Simulation aufzeichnen und wiedergeben kann. Dabei ist entscheidend, dass die Aufzeichnung die Dreidimensionalität erhält und der Blickwinkel bei der Wiedergabe somit frei gewählt werden kann. Wichtigste Bedingungen für den Erfolg der simulationsinternen Aufzeichnung sind sowohl eine kleine Dateigröße durch Kompression, als auch eine hohe Performanz von Aufzeichnung und Wiedergabe. Die Erfüllung dieser

Anforderungen wurde durch einen Test an einer repräsentativen Anlage verifiziert.

Zur Kompression der Daten wurde eine Kombination aus DEFLATE und DeltaEncoding Verfahren verwendet, welche durch Unterteilung des Datenstroms trotzdem das Springen zu willkürlichen Zeitpunkten während der Wiedergabe erlaubt. Durch diese Kombination konnte das durchschnittliche Datenvolumen von 1.098 MiB/min auf 3,4 MiB/min reduziert werden, was einem Kompressionsfaktor von 0,3 % entspricht. Erst dadurch wird die Speicherung von Langzeitaufnahmen auf handelsüblichen Festplatten ermöglicht.

Die Performanz von Aufzeichnung und Wiedergabe wurde durch Parallelisierung mit Hilfe von Microsofts Task Parallel Library verbessert.

Eine zweistufige Pipeline trennt bei der Aufzeichnung die Extraktion der Daten von der anschließenden Persistierung. Die Extraktionsstufe ist embarrassingly parallel und kann daher mit Hilfe des Fork-Join Modells parallelisiert werden. Die gesamte Zykluszeit der Simulation konnte dadurch um 47 % auf 10 ms reduziert werden. Diese liegt deutlich unter den, für 60 Hz notwendigen, 16 ms.

Zur Wiedergabe erfolgt der Ablauf in umgekehrter Reihenfolge der Aufzeichnung. Hier konnte die Zykluszeit um 52 % auf 15,3 ms, also ebenfalls unter 16 ms, reduziert werden.

Der interne Recorder wird das externe Aufnahmegerät ablösen, da die Aufzeichnung nicht auf einen Blickwinkel festgelegt ist, sondern bei der Wiedergabe individuell gesteuert werden kann.

Mit der Entwicklung von Recorder und Replayer wurde die Grundlage für ein um verschiedenste Daten erweiterbares Werkzeug für die 3D-Simulation gelegt. Der Aufnahmeumfang kann z. B. um Ein- und Ausgangsabbilder von Steuerungen ergänzt werden. Die dreidimensionale Aufzeichnung wird zukünftig auch außerhalb der VIBN in anderen Anwendungsfeldern des digitalen Zwillings eingesetzt werden. Mit Hilfe von VR kann beispielsweise bei virtuellen Schulungen die korrekte Bedienung von Maschinen anhand der Aufzeichnung demonstriert werden.

# Metadatenmanagement bei Data-Lake-Realisierungsansätzen

Dogus Tansel

Technische Hochschule Mittelhessen

Fachbereich MND  
Wilhelm-Leuschner-Straße 13  
61169 Friedberg  
E-Mail: [dogus.tansel@mnd.thm.de](mailto:dogus.tansel@mnd.thm.de)

Prof. Dr. Harald Ritz

Technische Hochschule Mittelhessen

Fachbereich MNI  
Wiesenstraße 14  
35390 Gießen  
E-Mail: [harald.ritz@mni.thm.de](mailto:harald.ritz@mni.thm.de)

## Kategorie

Bachelorarbeit

## Schlüsselwörter

Metadatenmanagement, Metadaten, MDMS, Data Lake, Data Factory, Metadaten-Repository, Metadatenpflege, Schema-on-Read

## Zusammenfassung

Im Zeitalter von Big Data werden täglich mehrere Billionen Daten weltweit erzeugt und zwischen unzähligen Systemen bewegt und verarbeitet. Sei es von industriellen Maschinen, AI-Software oder vom Menschen selbst erschaffen, so müssen diese enormen Datensätze gespeichert und gepflegt werden. Im Zuge der Digitalisierung ist es für Unternehmen umso wichtiger ausreichende Informationen über ihre Daten in Form von Metadaten zu erhalten, um Transparenz zu erzeugen, wie diese aufgebaut und verwendet werden. Diese gewaltigen Mengen an divergierenden Daten müssen in Systemen auf sorgfältiger Weise abgelegt und verwaltet werden, um einen ausreichenden Mehrwert aus den gewonnenen Informationen ziehen zu können.

Heutzutage fallen eine Vielfalt von möglichen Arten von Daten an, welche nicht mehr in klassischen Datenbanken bzw. Data-Warehouse-Systemen für Anwendungsfälle im Bereich Business Intelligence und Data Analytics gespeichert werden können, wie z.B. Bilder, Musik, Serverdaten u.v.m. Um solch ein breites Cluster an Daten abfangen und im jeweiligen Fall nutzen, sowie mithilfe von Metadaten in einem verwertbaren Rahmen verwalten zu können, bedarf es einer neuen und zeitgetreuen Alternative zu herkömmlichen Speicherverfahren. Ein Data Lake bietet diese Möglichkeit zum Verwalten und Analysieren heterogener, komplexer sowie umfangreicher Daten jeglicher Art, um die Realisierung eines Metadatenmanagementansatzes anzutreiben. Für dieses Unterfangen ist das Schema-on-Read-Verfahren in einem Data Lake zu realisieren. In einem klassischen Data Warehouse bildet die verwendete Datengrundlage ein Konstrukt an Informationen, das bereits zu Beginn des Ladeprozesses transformiert worden ist. Mit dem Schema-on-Read-Verfahren werden die verwendeten Daten erst zu Beginn einer Notwendigkeit in ihre entsprechende Form transformiert und somit dem

Anwender innerhalb des Data Lakes zur Verfügung gestellt.

Mit der vorliegenden Abschlussarbeit wird der Frage nachgegangen, welche Rolle das Metadatenmanagement in der heutigen Zeit einnimmt und wie dieses in einem modernen und zeitgetreuen Datenbankansatz zu realisieren ist. Hinsichtlich dieses Themas werden Implementierungsbeispiele für das Metadatenmanagement in Data Lakes mit strukturierten, semi-strukturierten und sog. unstrukturierten Daten realisiert, die Bedeutung von Metadatenmanagement-Konzepten und deren Technologien nähergebracht sowie die positiven und negativen Aspekte eines solchen Ansatzes dargestellt.

Als Ergebnis wurde deutlich, dass durch das Entstehen von Metadaten Unternehmen dazu getrieben werden ihre bisherigen Kenntnisse des Datenmanagements zu erweitern und sich im Spektrum einer metadatengetriebenen Architektur des Metadatenmanagements einzuordnen haben, um den Nutzern einer internen BI-Plattform im Rahmen von Self-Service- und Governance-Anforderungen gerecht zu werden.

Weiterhin ist mit der Implementierung eines Data-Lake- und Metadatenmanagement-Systems mit cloudbasierten Komponenten des Anbieters Azure und Google Cloud Products festzuhalten, dass das Metadatenmanagement eine zu bewältigende Herausforderung für Unternehmen und Industrie darstellt, um konsequente Datenhaltung über Systemgrenzen hinweg zu gewährleisten. Durch den Einsatz von ausgewählten IT-Werkzeugen, wie z.B. der Data Factory, die als cloudbasierter Datenintegrationsdienst die Pflege und das Verwalten großer Mengen von Informationen, insbesondere von Metadaten, ermöglicht, ist das Erfüllen der zentralen Ziele des Metadatenmanagements gewährleistet.

Diese Ziele sind zum einen die Minimierung der Aufwände beim Erstellen und im Betrieb von Informationssystemen mit großen Mengen an Daten und Metadaten und zum anderen das Erreichen eines weitreichenden Mehrwerts in der Informationsgewinnung für Entwickler und Endbenutzer.