Wirtschaftsinformatik 2023 Proceedings Wirtschaftsinformatik

10-9-2023

How Engineers perform Data Science Work: Designing Hybrid Roles

Amelie Lena Schmid
TU Dortmund, Germany, amelie.schmid@de.bosch.com

Manuel Wiesche
TU Dortmund, Germany, manuel.wiesche@tu-dortmund.de

How Engineers perform Data Science Work: Designing Hybrid Roles

Research Paper

Amelie Schmid[1,2] and Manuel Wiesche[1]

[1] TU Dortmund University, Dortmund, Germany
{amelie.schmid,manuel.wiesche}@tu-dortmind.de
[2] Robert Bosch GmbH, Reutlingen, Germany
amelie.schmid@de.bosch.com

Abstract. As of today, organizations are still struggling to derive consistent value from data science projects. The basic relevance of domain knowledge for data science work can be considered as common sense. Engineers, in particular, offer a unique view on emerging data science work, based on their critical role within traditional industries. As a constraint, current studies on data science work consider domain experts as rather passive and engineering-related studies are rare. To further explore these challenges, the present study analyses data science work of 30 engineers at an international automotive supplier. Investigating three cases, the evolvement of hybrid data science work can be derived, which combines two perspectives: engineering and data science. Thus, engineers actively incorporate the data science perspective, particularly when development activities involve minimal participation of data scientists. This contribution significantly enhances existing knowledge by demonstrating how engineers embrace the data scientists' perspective and perform hybrid data work.

Keywords: Data science work, data scientists, engineers, hybrid practice, transformation of work

1 Introduction

During the last years, data science projects have gained significant importance as an essential approach for organizations to leverage the vast amounts of available data. However, organizations often face major challenges with data science projects like long project cycles, never-ending pilots or even failed projects (Joshi et al. 2021). It results in the inability of extracting meaningful insights and creating consistent business value (Günther et al. 2017). One of the key issues lies in the existence of an expertise gap between data scientists and the domain experts, with their specialized knowledge in a specific domain (Kayabay et al. 2022). In this context, it is already acknowledged that a close collaboration between data scientists and domain experts is a success factor for

18th International Conference on Wirtschaftsinformatik,
September 2023, Paderborn, Germany

data science implementation as well as the performing of data science work, respectively (van den Broek et al. 2021). It can be explained by the fact that – in contrast – a forced technology implementation might generate a feeling of fear which leads to a rejecting attitude of the domain experts (Pachidi et al. 2021; Tams 2022). This holds in particular for highly educated groups like engineers, as these might react negatively to a new technology, if they feel attacked by it (Anthony 2018).

However, current studies regarding the collaboration of domain experts and data scientists consider the role of domain experts as rather passive, with only limited specific contribution towards the emerging data science work. Instead, domain experts are asked to simply provide their domain knowledge and share their know-how in a way that the data scientists are able to make use of it in the further development process (Lebovitz et al. 2021; Park et al. 2021; Sambasivan and Veeraraghavan 2022). As a consequence, the resulting data science work is predominantly performed by the data scientist.

Digging into deep, the current insights regarding data science work of domain experts are predominantly grounded on studies within business-related areas like in HR (van den Broek et al. 2021), sales (Pachidi et al. 2021) or consulting (Strich et al. 2021). However, the professional group of engineers is of particular high relevance, as they represent a key subject for the ongoing digital transformation within traditional industries, while facing at the same time fundamental changes with respect to their existing tasks and role (Azmi et al. 2018; Muhuri et al. 2019). Furthermore, the special characteristic of engineers within organizations, e.g. being conceptually driven as well as their critical contribution towards product development and innovation require to take a differentiated look at this group of domain experts (Bechky 2003; Menzel et al. 2007).

Considering the relevance of domain knowledge for successful data science work (van Giffen and Ludwig 2023), the present work aims to expand the existing perspective of data scientists by focusing on the unique role of engineers. Thus, the following research question will be answered: *How do engineers perform data science work?*

The data science work of engineers is analyzed based on 30 semi-structured interviews within one of the biggest automotive supplier. The evolvement of a hybrid work practice which is predominantly performed by the domain experts is discussed. Finally, the interconnection of the given work practices of engineers and the incorporation of new work practices of data scientists is discussed.

2 Theoretical Background

2.1 Data Science Work & Engineering Work

Recently, data science gained increasing attention within organizations, as it can be used to gather new insights and create additional value from data (Günther et al. 2017). So far, no consistent designation for data science and its work practices exists. The used terms and definitions within information systems (IS) research range from naming a specific method like the usage of AI (Strich et al. 2021), big data, (Provost and Fawcett 2013), machine learning (Meyer et al. 2014) or robotics (Willcocks 2020). Moreover, these specific fields are summarized and only generic terms like data-driven work

(Brynjolfsson et al. 2011) or data work (Parmiggiani et al. 2022) are applied. In order to frame the scope of our study, we refer to data science as the practice of analyzing large-scale data to get a glimpse on decision-making and taking actions (Abbasi et al. 2016; Crisan et al. 2021; Passi and Sengers 2020). Basically, the work of a data scientist is characterized by the detection of correlations and patterns within a defined data set. The overall goal of their work is to reveal insights based on data, even though these might not be generalizable, due to the missing direct link to the underlying reasons and causalities (Provost and Fawcett 2013). Thus, *data science work* is driven by rationality, objectivity and neutrality. The behavior of data scientists is evidenced-based, having the goal to diminish human bias (Agarwal and Dhar 2014; Davenport 2018; Jones 2019; Power et al. 2019).

In contrast, an exact understanding of the specific problem is of great importance for *engineering work*. That means that engineers apply their specific domain knowledge to generate solutions and use e.g. physical approaches to generalize their results (Anderson et al. 2010; Sheppard et al. 2006). Moreover, the behavior of engineers is rather conceptual, having the goal of designing e.g. new products (Bechky 2003). During the development process, engineers apply different strategies in order to guide others towards a particular meaning. Thus, engineers might want to utilize data to support an intended preference (Barley et al. 2012). It can be concluded that the basic approach and goal of data science work differs from engineering work to a large extent.

2.2 Transformation of Knowledge Work

It is already well known that work practices change based on increasing technological usage (Vaast and Walsham 2005). Also data science offers tremendous new opportunities and creates different demands on the work of any domain expert (Leitner-Hanetseder et al. 2021). In the present study, work will be defined as “the use of human, informational, physical, and other resources to produce products/services” (Alter 2013). Especially in the knowledge work context, data science has a major impact (Lebovitz et al. 2021).

In general, it is usually assumed that the transformation of knowledge work does not originate from the domain experts themselves. Klicken oder tippen Sie hier, um Text einzugeben.Instead, work transformation might be triggered by a forced IT system implementation (Soh et al. 2003; Vaast and Walsham 2005) or activated by an external market change (Nelson and Irwin 2014). In both cases numerous challenges show up for the domain experts. The negative effects of a forced work transformation are already well known: research on forced IT system implementations gave evidence on issues like decreasing job satisfaction due to less job control on the side of the employee (Bala and Venkatesh 2013) or challenges regarding workarounds and unintended use (Boudreau and Robey 2005). Prior studies have additionally focused on work transformation, where external market change caused an adaptation of the working mode without any alternative. For example, Nelson and Irwin (2014) illustrate how librarians had to change their work as an reaction towards the emerging internet. The so-called “paradox of expertise“ lead to the fact that they refrained from integrating the new technology into their working mode by themselves, although the capabilities were available.

Moreover, current studies consider data science applications which are predominantly developed by data scientists and not by domain experts (Strich et al. 2021; van den Broek et al. 2021; Waardenburg et al. 2022; Zhang et al. 2020). One reason is the lacking capability of domain experts to apply data science, especially in business professions with limited data-related understanding (Oberländer et al. 2020). Instead, data scientists are requested to closely align their developing activities with the domain experts in order to integrate the specific knowledge into the technology, as domain-specific knowledge is a crucial success factor for data science work (Lebovitz et al. 2021; van den Broek et al. 2021). However, during the collaboration of domain experts and data scientists, additional challenging effects such as limited or even lacking communication between the two groups occur. Maintaining a common motivation and consistently following joint goals represent further issues which hinder successful collaboration and data science work (Mao et al. 2019). As a consequence, the domain experts might not be able to interpret and evaluate the outcome of the developed algorithms, as the internal logic is unknown (Zhang et al. 2020).

All the mentioned challenges might be tackled by granting domain experts an active role in the development of data science. Therefore, it is required to overcome the knowledge boundary between data scientists and domain experts (Waardenburg et al. 2022). In the IS context, the active participation in workforce advice networks during an IT system implementation serve as a first example for that. Such a high employee interaction at the workplace during the implementation phase has a positive influence on job performance (Sykes et al. 2014). However, most studies consider the formation of data science work from data scientist's point of view and leave domain experts with the passive attitude of providing and sharing domain knowledge only (Park et al. 2021).

To conclude, it remains unclear how domain experts perform data science work, independently form the strong influence of data scientists. This holds especially if the development of data science application takes place by themselves, e.g. due to the given data understanding of engineering. Thus, it is necessary to put a focus on data science work based on the application of data science by engineers.

3 Methodology

In this research, we chose a qualitative research approach to examine data science work, performed by the domain expert instead of the data scientist. Specifically, we analyzed data science work by engineers within three cases. We find the qualitative approach particularly helpful, as it allows to adapt a research process towards the evolving research context in an iterative manner (Charmaz 2014; Corbin and Strauss 2015). Thus, it enables the answering of "how" questions (Langley 1999; van de Ven and Huber 1990).

3.1 Data Collection and Immersion into the Field

The analysis is based on a study within one of the biggest suppliers within the automotive industry, which is developing and manufacturing semiconductors and sensors as well as electronic control units (ECUs) at 40 plants and locations worldwide. We choose the automotive industry because it is a strategically relevant showcase for the complex digitalization process of traditional companies and the related engineers. The analysis includes frequent visits to the field over the course of 12 months (Walsham 1995). Thereby, we have involved ourselves deeply into the context in order to truly understand the ecosystem (Johns 2006).

We chose three cases within the organization, as the analytical conclusions of multiple-case designs are generally more powerful and robust compared to single-case designs (Miles et al. 2020; Yin 2010). Case similarity and case variability served as guiding principles for the selection of the cases (Kirsch 1997; Orlikowski 1993). To ensure case similarity, we chose the organizational area related to the production of electronic control units (ECUs) as the overall context of the study. The utilization of data science plays a major role in this entire area, as it had been increasingly applied at different points along the entire ECU value stream. Thus, the environmental conditions and the data availability can be considered as a common ground. However, the cases varied in involved engineers, departments, functions, specific tasks, stakeholders and localization.

Table 1. Overview of the three cases

Characteristics	Case 1: Supplier	Case 2: R&D	Case 3: Manufacturing
Traditional work design	Manual claim management of each individual material; shipment of the broken part to the supplier; discussion with supplier ex-post, based on the physical product.	Product development by using domain knowledge; decision-making based on experience of the employees, as well as analytical models, basic statistics / simulation.	Manufacturing problems are solved by reacting to emerging issues; evaluation of the situation by an operator at the line; non-systematic quality management.
Work design with data science application	Claim management based on manufacturing data from production lines; no shipment of broken parts; exchange of data.	Data-driven product development; decision making based on data models by using manufacturing data as well as field data.	Systemized problem analysis and solving using manufacturing data and visualization; data-driven quality assurance.

Table 1 represents an overview of the three selected cases: (1) Supplier, (2) R&D and (3) Manufacturing. Each case has been selected based on a specific purpose within the overall research (Yin 2010). Case 1 was chosen as it involves engineers who closely interact with external suppliers. The engineers selected from the R&D departments of case 2 bring in the perspective of highly educated employees, having scientific backgrounds like physics, chemistry and material science. Case 3 serves as a good field of investigation, as it is related to the core process of the company and the involved engineers need to meet the highest quality standards.

In order to generate a profound picture we sampled 30 engineers across all three cases from different hierarchical levels, a varied degree of data science knowledge and working experience as well as a mixture of backgrounds and work roles (Patton 2009). Data was collected using semi-structured interviews (Miles et al. 2020). Within these interviews our primary goal was to elaborate on the emerging data science work practice. For that purpose, in a first step, we asked the participants to describe their current tasks in detail. Subsequently, this was followed by the elaboration of the new work practices with a focus on recent changes based on increasing data availability. The preliminary interview guideline was further optimized by follow-up questions as data collection progressed (Miles et al. 2020). The duration of the interviews was ~54 minutes.

Even though interviews were the primary data source, we collected further qualitative data to enhance the fundamental understanding and to receive a clearer picture of the phenomenon (Walsham 1995). The data included different documentations of the reported data science use cases which have been created for internal management reviews. In addition, we participated two times at a 2-day workshop with representatives of the relevant functions, where the current data science projects have been elaborated and discussed.

3.2 Data Analysis

To obtain a profound overview of the key messages during the data collection process, we applied memoing and wrote a short summary with the most relevant statements after each interview (Glaser 1978). These the memos helped to further remember all relevant data during the data collection process. Consecutively, we were able to get back to the raw data during the data analysis. We transcribed all interviews, resulting in 482 pages of qualitative data, which serves as the primary data of this study.

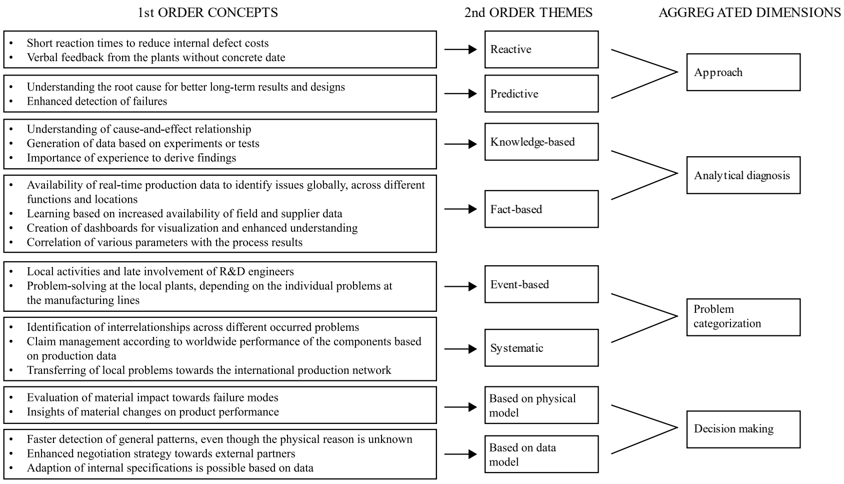

Figure 1. Illustration of the coding process

The data analysis followed an iterative procedure. A qualitative data analysis software (MAXQDA) was used to gather, organize, and analyze the data in a clear and transparent manner. We coded all interview transcripts systematically, applying grounded theory practice (Corbin and Strauss 2015; Sarker et al. 2018). In a first step, open coding was used to assign appropriate labels to the interview passages which summarized the key message in a short phrase (Miles et al. 2020). Here, we identified the general work characteristics in the context of data science application. During the data analysis process, we realized the persistence of the former engineering work practices to a large extent. That is why we used axial coding for the identification of connections between the subcodes (Strauss 1978). We were able to map the former engineering work practices as well as the new data science work practices to each work element. Finally, we applied selective coding to further develop and combine the dimensions, resulting in the final consolidation (Figure 1) (Glaser 1978). The whole coding process followed an iterative procedure through several rounds of reviewing and rewriting (Miles et al. 2020).

4 Results

The present work is targeting on a deeper understanding of data science work, performed by engineers. As a result, a hybrid data science practice is presented, illustrating that engineers combine (i) elements from their established work practices and (ii) the new work practices of data scientists. The characteristics of the hybrid role can be detailed by the following four relevant work elements: (i) approach, (ii) analytical diagnosis, (iii) problem categorization and (iv) decision-making (Figure 2). Each of these categories and the respective characteristics of the two roles will be elaborated in the following.

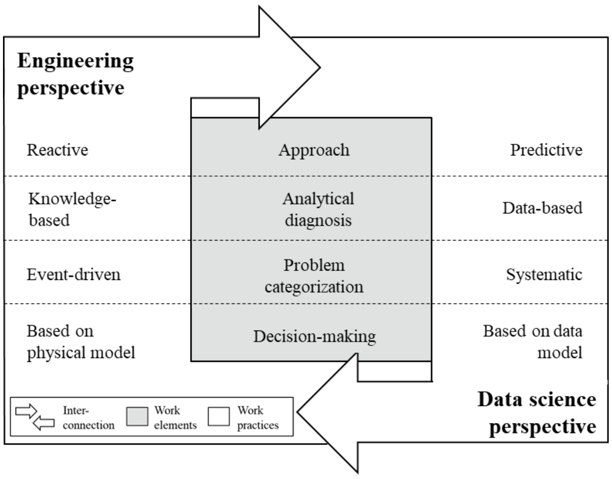

Figure 2. Hybrid practice of data science work

Combining a reactive and a predictive approach. Both the reactive and predictive aspects are highly relevant for data science work of engineers within all three cases. On the one hand side, the *reactive* work practices are well established since engineers need to support in problem solving during the manufacturing. Short reaction times are needed, as internal errors cause a high amount of internal cost. The collaboration between R&D and manufacturing requires to stay reactive at specific circumstances:

"The major pain point is that we get a lot of verbal feedbacks from the plants regarding their problems. They are saying: […] last night we had 10 pieces of failures, please support us." (#28)

Nevertheless, the identification of such incidences serves as an important baseline to learn for the future. Here, the application of data science is essential. Thus, the reactive working approach is strengthened by the *predictive* perspective. The proactive usage of data tremendously increases both the speed of product development and the failure detection. In addition, the adaptation of current operational processes can improve the quality of future work. As a result, plants can trigger improvements, such as machine settings with reference to identified errors. All in all, this targets to prevent reoccurring issues based on the same root cause.

In addition, by means of the explorative identification of similar error structures future issues can be avoided. Such an early warning systems helps to keep the operations running within defined control limits, as potential issues are detected in advance. The roll-out and the proactive integration of the developed solution into future products by the workers form R&D ensure a reduction or even elimination of specific failures. Such a mindset of forward-thinking in R&D helps to introduce fundamentally new ideas:

"With much more detail we can clearly identify the root cause in comparison to a hypothesis […]. And now we can understand: 'Ok, if I do this, it will make this happen and in the end of the line we will have a good result, because I understand what is happening in between.' […] And this also allows the team (…) to make better designs from scratch." (#27)

Analytical diagnosis based on knowledge and data. Engineers have acquired profound domain knowledge during their studies and several years of working experience complement the knowledge level to a high extend. Hence, *knowledge-based* diagnosis is of great importance for certain operations. One respondent highlights the benefits of knowledge application, like an intense and autonomous analysis of any problem. For that purpose, a deeper involvement into the field is necessary to deal with the situation and to deduce a solution simply based on domain knowledge.

"The bottom line of my work is to build up knowledge. And I believe that this will remain the same and that you will continue to need the expertise, the experience." (#22)

However, in some situations the exact identification of failures and error patterns is complex. This can even lead to an entire lack of understanding on relevant topics or occurring issues. To tackle that data science is strongly needed to carry out a complementary *data-based* analysis. Facts and figures, enhance the problem diagnosis in various ways. In a first step, a detailed visualization of the manufacturing data can help to get a general feeling for the respective circumstances:

"Even the development who is very, very far away from production can check and see that: 'ok, we have a problem at this and this station and that press fit force has been increased for the past two hours.' […] And then […] I see: 'ok, the force increase is caused by this supplier batch of the pins.' And this is making it much easier to detect the root cause of the problem." (#26)

This is valid if visualizations of relevant information are continuously available and always up to date. Using real-time data enables a constant tracking of the processes within manufacturing. In this way, the influence and effects of individual activities of workers form R&D or manufacturing can be identified immediately:

"But for such regular queries I created a Tableau dashboard. [...] There you see nicely: after we made the change here, that we got a nice improvement. From B/C o mainly A, but also a little bit of B still there. And to track that effectiveness, I just quickly made a dashboard." (#13)

Considering case 2, it can be concluded that the application of data science even leads to a systematic verification of already existing knowledge-based assumptions. In this context, the scope of the current work within R&D can be expanded since it is now possible to investigate relevant aspects by means of available data. In the past, models were built up and, in some cases, no clear explanation could be found. Today and in the future data science application offers the possibility to check and calibrate these already existing models based on the data. Thus, so far open questions e.g. regarding the maximum storage time of printed circuit boards (PCBs) can now finally be answered:

"What is the right storage time of PCBs [printed circuit boards]? That is very difficult to check on the development side. We try to anticipate the ageing process, like calculating how much oxide growth I have. But in the end, you can also say: 'Hey, the truth is in the field, right'? Let's see how old the PCBs are, whether we see a correlation, for example, that PCBs which are one year old have a higher defect rate than completely fresh PCBs." (#30)

Using event-driven and systematic problem categorization. Our data reveal that the problem categorization is another important component, which is a relevant data science work practice. As already indicated, in the context of problem solving the established work practice of engineers are rather event-driven. Case 3 shows that single, specific events must be analyzed in order define a solution for the specific problem. In particular, this holds if the occurring issue needs to be solved fast. Timing is another crucial factor, which explains an apparently randomized way of working:

"The difficulty is: I always have the problems in the plants at different point in times. That's tricky, because today plant A calls, tomorrow plant B calls." (#09)

This event-driven way of working results in work being categorized as single issues. Local on-site teams are set up and start to work jointly on the specific issue. Such an event-driven procedure has to be guided by a systematic, all-time solution of the problem. As one key aspect, data science shifts the perspective and generates the required holistic view. It needs to be taken care that single events at specific local plants are excluded from the data, not to endanger that those local effects overlay the entire picture. Such local events are among others incorrect machine settings or local environmental conditions which are not design-related or linked to the specific component.

As an essence, the systematic problem categorization focuses on the global perspective and distinguishes between real global pain points and specific location-dependent effects. The availability of manufacturing data offers the opportunity to detect the root cause of a problem at any line in a structured way. The priority of local effects will be decreased, and the focus is set on global incidences which can be solved with much more effectiveness. In this way, it is possible to identify interrelationships of different order which can cause further possible occurring issues. Thus, it changes the problem categorization with respect to the claim management procedure fundamentally. Even the work with external partners is changing, as the systematic usage of data provides a new basis for collaboration. This can be illustrated by the following example which describes how the work practice with external suppliers is adapted (case 1):

"It is possible to make a claim of the performance of the components based on production data. This means that not only individual cases can be claimed, but also the global performance of different plants as well as differences in the performance of different suppliers." (#29)

The systematic problem categorization can also help to implement continuous improvement methods within the international production network. In this way, problems are not only solved at that single point in time, but are also faster detected in case of similar occurrences at the different locations:

"We share all information in an international lesson learned network so that everyone in the organization can make use of a holistic approach and transfer it to his cases." (#03)

Decision-making grounding on physical model and data model. In the daily work, engineers have to make countless decisions, using physical and data model as a combination. Decisions based on physical models do not necessarily require large amount of data for verification. Instead, engineers build a basis for the decision through the application of models which are referring to physical laws. These physical models are available for every engineering topic, such as mechanics, hardware or manufacturing development. This serves as an important baseline for decision-making in the R&D area, like the selection of the appropriate material in the product design:

"If I change the material from copper to iron, I would like to know if this has an influence on my failure mode, e.g. the temperature cycling performance. My physical model tells me that stress and strain can cause cracks in the solder joint [...] and that iron promotes this more than copper, as the coefficient of thermal expansion is more critical. I have not recorded a single data point; I have derived analytically from physical models that this is going in the wrong direction." (#02)

The analytical derivation of the cause-and-effect relationship based on physics of failure enables this verification, without the application of data science. In contrast, data and thus the application of data science is the fundamental prerequisite for deriving decisions based on data models. This holds for decisions within all three cases like the selection of a supplier (case 1), product development decisions (case 2) or decisions regarding the right way to solve occurring problems (case 3). By using the data model approach, it is possible to identify issues which cannot be explained solely by a physical model. Instead, it is possible to see further indications which might cause an issue. One example is the correlation between the results and a certain supplier, which enables insights into their respective performance and creates a better basis for negotiation:

"Based on data, we can try to change the specification for the PCB [printed circuit board] shrinkage, for the PCB (reflow) cycles. It is hard work but based on our statistic regression we see that the range of the current defined specification is too big, and we can reduce it." (#28)

However, the accuracy of decision-making based on data models is strongly dependent on their granularity and level of detail. It is important to have a comprehensive description, leading in the best case to a digital twin that contains any relevant element of the real world. At the final stage, the algorithm is even able to develop and recognize its own patterns with huge amount of data. It is important to mention that the examples of the decision-making aspects illustrate how the two poles of the physical model and data model approach interact with each other. At the first glance, decisions by data models based on probabilities can be taken, which do not require any domain knowledge. However, the physical reason behind these failures is still lacking and unclear. Digging deeper into the topic shows that results from the data model are not enough and hence, have to be explained by a physical reason:

“Of course, based on the data we would only see that the workpiece carrier is causing the issue and when we check it physically, we know this is being caused by the loose screw.” (#26)

Combining engineering and data science work practices towards an interconnected hybrid practice. To sum up, engineer perform data science work by integrating data science practices into their existing role. Depending on the phenomenon, it can be required to apply the sophisticated engineering work practices. However, the merger with data science leads to the creation of the new hybrid practice. As a result, the feedback loop can be closed by the interconnection of both work practices:

“Particularly regarding our manufacturing, we can close this circle, because at the end of the day, we always view our work as CIP [continuous improvement process]. And if we include the feedback, we can look at how we can improve.” (#20)

Finally, it can be recognized that both perspectives are strongly interwoven and in constant interaction. The traditional approach serves even as a baseline for fueling the data science perspective. For example, the generated data from event-driven issues within case 3 are an important prerequisite for further enhancing the R&D process:

“We have developed an IDC [internal defects cost] reduction cycle. Here we have both, the backwards integration to existing products and forward integration for new products.” (#02)

5 Discussion

The scope of the present study is to understand the design of data science work by engineers. The integration of data science into the existing work of the domain expert leads to the evolvement of hybrid data science work. Hence, the content of the current literature is expanded in two ways. First, it is introduced that the domain experts are the drivers of their data science work. Second, we show that engineering domain experts embed data science into their existing work, resulting in the hybrid practice.

5.1 Theoretical and Practical Contributions

For data scientists, it is already well known that they utilize data to construct new ways of working (Muller et al. 2019). In contrast, domain experts were so far considered as being passive in the sense of performing data science work (Jung et al. 2022; van den Broek et al. 2021). In the present work, we broaden this limited view and illustrate a new work setting which is created by the engineers themselves. Consequently, data science work is performed by domain experts which leads to the fact that engineers question their existing boarder criteria. Moreover, the workers even calibrate and verify their established work results by means of data science.

Among others, this can be emphasized e.g. in the sense of decision-making. Here, the solely application of data science, results in uncertainty and missing underlying explanations (Lebovitz 2019; Lebovitz et al. 2022; Provost and Fawcett 2013). Nonetheless, work practices of engineering like the understanding of physical causalities remain relevant. As a result, the combination of engineering and data science allows profound understanding, which provides further evidence on the relevance of the hybrid work practices. Prior studies already emphasized the importance of creating hybrid systems in the context of machine learning, as AI can only partially solve problems (Zhang et al. 2020). Within this work, we illustrate such a hybrid approach in the context of data science and provide for the first time a detailed explanation of the demanded hybrid practice. In addition, it serves as a detailed example for an emergent evolvement of digital transformation, as domain experts continuously realize value from the adjustment of their work (Koch et al. 2021).

The findings further extent recent studies investigating the impact of technology on work (van den Broek et al. 2021; Willcocks 2020). As an example, Strich et al. (2021) show that the old tasks of consultants were eliminated by an AI system. However, we demonstrate that domain experts do not discard the existing work practices. Instead, data science complements the established work, and it becomes part of the existing job. In this way, we contribute to research which recognizes that technology will rather modify than completely replace existing jobs (Bughin 2018; Huysman 2020) as well as research on the evolvement of an “IT identity“ (Carter and Grover 2015). We extent this body of knowledge by illustrating two underlying success criteria concerning domain experts: First, the stronger contribution compared to the data scientists and second, the capability to perform new data science work by themselves.

As an outcome, we derive contributions to practice from this study. First, it provides a more differentiated view on the establishment of new data roles within organizations. Major aspects of domain experts should be kept to a large extend, instead of pushing a hard transformation from classical engineers towards data scientists. This procedure helps to overcome obstacles and to convince additional domain experts to go into the same direction. Second, the results of this study reveal that the commitment towards data science application plays a crucial role for the successful implementation. It is therefore important to share the success story within the company to influence others.

5.2 Limitations and Future Research

The present study is subject to some limitations. The qualitative approach and the selective sample size results in restrictions on the overall generalizability of the results, as the data collection is limited to the domain-specific context with a focus on the automotive industry. An effort was spent to counteract this obstacle by selecting four different cases within the overall context. However, further research could seek to widen this perspective and compare the results to domain experts from different industries.

Finally, our work focusses on the specific characteristics of the created hybrid data science work, while disregarding the respective transformation process towards that hybrid practice. It might be interesting to see how the transformation path works and whether the well-known “Model of Institutionalizing“ (Reay, T., Golden-Biddle, K., & Germann, K. 2006) also holds in the data science context.

References

Abbasi, A., Sarker, S., and Chiang, R. 2016. “Big Data Research in Information Systems: Toward an Inclusive Research Agenda,” *Journal of the Association for Information Systems* (17:2), pp. 1-32 (doi: 10.17705/1jais.00423).

Agarwal, R., and Dhar, V. 2014. “Editorial — Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research,” *Information Systems Research* (25:3), pp. 443-448 (doi: 10.1287/isre.2014.0546).

Alter, S. 2013. “Work System Theory: Overview of Core Concepts, Extensions, and Challenges for the Future,” *Journal of the Association for Information Systems* (14:2), pp. 72-121 (doi: 10.17705/1jais.00323).

Anderson, K. J. B., Courter, S. S., McGlamery, T., Nathans-Kelly, T. M., and Nicometo, C. G. 2010. “Understanding engineering work and identity: a cross-case analysis of engineers within six firms,” *Engineering Studies* (2:3), pp. 153-174 (doi: 10.1080/19378629.2010.519772).

Anthony, C. 2018. “To Question or Accept? How Status Differences Influence Responses to New Epistemic Technologies in Knowledge Work,” *Academy of Management Review* (43:4), pp. 661-679 (doi: 10.5465/amr.2016.0334).

Azmi, A., Kamin, Y., Noordin, M., and Nasir, A. 2018. “Towards Industrial Revolution 4.0: Employers' Expectations on Fresh Engineering Graduates,” *International Journal of Engineering and Technology* (7:28), pp. 267-272.

Bala, H., and Venkatesh, V. 2013. “Changes in Employees' Job Characteristics During an Enterprise System Implementation: A Latent Growth Modeling Perspective,” *MIS Quarterly* (37:4), pp. 1113-1140 (doi: 10.25300/MISQ/2013/37.4.06).

Barley, W. C., Leonardi, P. M., and Bailey, D. E. 2012. “Engineering Objects for Collaboration: Strategies of Ambiguity and Clarity at Knowledge Boundaries,” *Human Communication Research* (38:3), pp. 280-308 (doi: 10.1111/j.1468-2958.2012.01430.x).

Bechky, B. A. 2003. “Sharing Meaning Across Occupational Communities: The Transformation of Understanding on a Production Floor,” *Organization Science* (14:3), pp. 312-330 (doi: 10.1287/ORSC.14.3.312.15162).

Boudreau, M.-C., and Robey, D. 2005. “Enacting Integrated Information Technology: A Human Agency Perspective,” *Organization Science* (16:1), pp. 3-18 (doi: 10.1287/orsc.1040.0103).

Brynjolfsson, E., Hitt, L. M., and Kim, H. H. 2011. “Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?” in *Proceedings of the 32nd International Conference on Information Systems*, Dennis F. Galletta and Ting-Peng Liang (eds.), Shanghai, China. December 4-7 (doi: 10.2139/ssrn.1819486).

Bughin, J. 2018. “Why AI isn’t the Death of Jobs,” *MIT Sloan Management Review*.

Carter, M., and Grover, V. 2015. “Me, My Self, and I(T): Conceptualizing Information Technology Identity and its Implications,” *MIS Quarterly* (39:4), pp. 931-957 (doi: 10.25300/MISQ/2015/39.4.9).

Corbin, J. M., and Strauss, A. L. 2015. *Basics of Qualitative Research*: *Techniques and Procedures for Developing Grounded Theory*, Los Angeles: SAGE.

Crisan, A., Fiore-Gartland, B., and Tory, M. 2021. "Passing the Data Baton: A Retrospective Analysis on Data Science Work and Workers," *IEEE transactions on visualization and computer graphics* (27:2), pp. 1860-1870 (doi: 10.1109/TVCG.2020.3030340).

Croft, C., Currie, G., and Lockettt, A. 2015. "Broken 'Two-Way-Windows'? An Exploration of Professional Hybrids," *Public Administration* (93:2), pp. 380-394 (doi: 10.1111/padm.12115).

Davenport, T. H. 2018. *The AI Advantage*, The MIT Press.

Denzin, N. K. 2017. *Sociological Methods*, Routledge.

Glaser, B. G. 1978. *Theoretical sensitivity*, Mill Valley, Calif.: Soc. Pr.

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., and Feldberg, F. 2017. "Debating Big Data: A Literature Review on Realizing Value from Big Data," *The Journal of Strategic Information Systems* (26:3), pp. 191-209 (doi: 10.1016/j.jsis.2017.07.003).

Huysman, M. 2020. "Information Systems Research on Artificial Intelligence and Work: A Commentary on "Robo-Apocalypse cancelled? Reframing the automation and future of work debate"," *Journal of Information Technology* (35:4), pp. 307-309 (doi: 10.1177/0268396220926511).

Iyer, L., Church, M., and and Emery, D. 2018. "An In-depth Analysis of Careers in Data Science: A K-Means Clustering Approach," in *Proceedings of the 24th Americas' Conference on Information Systems*, M. Garfield (ed.), New Orleans, LA, USA. August 16-18.

Johns, G. 2006. "The Essential Impact of Context on Organizational Behavior," *Academy of Management Review* (31:2), pp. 386-408 (doi: 10.5465/AMR.2006.20208687).

Jones, M. 2019. "What we talk about when we talk about (Big) Data," *The Journal of Strategic Information Systems* (28:1), pp. 3-16 (doi: 10.1016/j.jsis.2018.10.005).

Joshi, M., Su, N., Austin, R., and Sundaram, A. 2021. "Why So Many Data Science Projects Fail to Deliver," available at https://sloanreview.mit.edu/article/why-so-many-data-science-projects-fail-to-deliver/.

Jung, J. Y., Steinberger, T., King, J. L., and Ackerman, M. S. 2022. "How Domain Experts Work with Data: Situating Data Science in the Practices and Settings of Craftwork," *Proceedings of the ACM on Human-Computer Interaction* (6:CSCW1), pp. 1-29 (doi: 10.1145/3512905).

Jussupow, E., Spohrer, K., Heinzl, A., and Gawlitza, J. 2021. "Augmenting Medical Diagnosis Decisions? An Investigation into Physicians' Decision-Making Process with Artificial Intelligence," *Information Systems Research* (32:3), pp. 713-735 (doi: 10.1287/isre.2020.0980).

Kayabay, K., Gökalp, M. O., Gökalp, E., Erhan Eren, P., and Koçyiğit, A. 2022. "Data Science Roadmapping: An Architectural Framework for Facilitating Transformation towards a Data-driven Organization," *Technological Forecasting and Social Change* (174), p. 121264 (doi: 10.1016/j.techfore.2021.121264).

Kim, M., Zimmermann, T., DeLine, R., and Begel, A. 2018. "Data Scientists in Software Teams: State of the Art and Challenges," *IEEE Transactions on Software Engineering* (44:11), pp. 1024-1038 (doi: 10.1109/TSE.2017.2754374).

Kirsch, L. S. 1997. “Portfolios of Control Modes and IS Project Management,” *Information Systems Research* (8:3), pp. 215-239 (doi: 10.1287/isre.8.3.215).

Koch, H., Chipidza, W., and Kayworth, T. R. 2021. “Realizing value from shadow analytics: A case study,” *The Journal of Strategic Information Systems* (30:2), p. 101668 (doi: 10.1016/J.JSIS.2021.101668).

Langley, A. 1999. “Strategies for Theorizing from Process Data,” *Academy of Management Review* (24:4), p. 691 (doi: 10.2307/259349).

Lebovitz, S. 2019. “Diagnostic Doubt and Artificial Intelligence: An Inductive Field Study of Radiology Work,” in *Proceedings of the 40th International Conference on Information Systems*, H. Krcmar and J. Fedorowicz (eds.), Munich.

Lebovitz, S., Levina, N., and Lifshitz-Assa, H. 2021. “Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What,” *MIS Quarterly* (45:3), pp. 1501-1526 (doi: 10.25300/MISQ/2021/16564).

Lebovitz, S., Lifshitz-Assaf, H., and Levina, N. 2022. “To Engage or Not to Engage with AI for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis,” *Organization Science* (33:1), pp. 126-148 (doi: 10.1287/orsc.2021.1549).

Leitner-Hanetseder, S., Lehner, O. M., Eisl, C., and Forstenlechner, C. 2021. “A Profession in Transition: Actors, Tasks and Roles in AI-based Accounting,” *Journal of Applied Accounting Research* (22:3), pp. 539-556 (doi: 10.1108/JAAR-10-2020-0201).

Mao, Y., Wang, D., Muller, M., Varshney, K. R., Baldini, I., Dugan, C., and Mojsilović, A. 2019. “How Data Scientists Work Together With Domain Experts in Scientific Collaborations,” *Proceedings of the ACM on Human-Computer Interaction* (3:GROUP), pp. 1-23 (doi: 10.1145/3361118).

Menzel, H. C., Aaltio, I., and Ulijn, J. M. 2007. “On the way to creativity: Engineers as intrapreneurs in organizations,” *Technovation* (27:12), pp. 732-743 (doi: 10.1016/j.technovation.2007.05.004).

Meyer, G., Adomavicius, G., Johnson, P. E., Elidrisi, M., Rush, W. A., Sperl-Hillen, J. M., and O'Connor, P. J. 2014. “A Machine Learning Approach to Improving Dynamic Decision Making,” *Information Systems Research* (25:2), pp. 239-263 (doi: 10.1287/isre.2014.0513).

Michalczyk, S., Nadj, M., Maedche, A., and Gröger, C. 2021. “Demystifying Job Roles in Data Science: A Text Mining Approach,” in *ICIS 2021 Proceedings*, M. Limayem, R. EL Amrani and F. Rowe (eds.), In Marrakech, Morocco. 14., p. 115.

Miles, M. B., Huberman, A. M., and Saldaña, J. 2020. *Qualitative data analysis: A methods sourcebook*, Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE.

Muhuri, P. K., Shukla, A. K., and Abraham, A. 2019. “Industry 4.0: A bibliometric analysis and detailed overview,” *Engineering Applications of Artificial Intelligence* (78), pp. 218-235 (doi: 10.1016/j.engappai.2018.11.007).

Muller, M., Lange, I., Wang, D., Piorkowski, D., Tsay, J., Liao, Q. V., Dugan, C., and Erickson, T. 2019. “How Data Science Workers Work with Data,” *CHI '19: CHI*

Conference on Human Factors in Computing Systems, pp. 1-15 (doi: 10.1145/3290605.3300356).
Nelson, A. J., and Irwin, J. 2014. "“Defining What We Do—All Over Again”: Occupational Identity, Technological Change, and the Librarian/Internet-Search Relationship," *Academy of Management Journal* (57:3), pp. 892-928 (doi: 10.5465/amj.2012.0201).
Oberländer, M., Beinicke, A., and Bipp, T. 2020. "Digital competencies: A review of the literature and applications in the workplace," *Computers & Education* (146), p. 103752 (doi: 10.1016/j.compedu.2019.103752).
Orlikowski, W. J. 1993. "CASE Tools as Organizational Change: Investigating Incremental and Radical Changes in Systems Development," *MIS Quarterly* (17:3), p. 309 (doi: 10.2307/249774).
Pachidi, S., Berends, H., Faraj, S., and Huysman, M. 2021. "Make Way for the Algorithms: Symbolic Actions and Change in a Regime of Knowing," *Organization Science* (32:1), pp. 18-41 (doi: 10.1287/orsc.2020.1377).
Park, S., Wang, A. Y., Kawas, B., Liao, Q. V., Piorkowski, D., and Danilevsky, M. 2021. "Facilitating Knowledge Sharing from Domain Experts to Data Scientists for Building NLP Models," in *26th International Conference on Intelligent User Interfaces*, T. Hammond, K. Verbert, D. Parra, B. Knijnenburg, J. O'Donovan and P. Teale (eds.), College Station TX USA. 14 04 2021 17 04 2021, New York, NY, USA: ACM, pp. 585-596 (doi: 10.1145/3397481.3450637).
Parmiggiani, E., Østerlie, T., and Almklov, P. G. 2022. "In the Backrooms of Data Science," *Journal of the Association for Information Systems* (23:1), pp. 139-164 (doi: 10.17705/1jais.00718).
Passi, S., and Sengers, P. 2020. "Making Data Science Systems Work," *Big Data & Society* (7:2), 205395172093960 (doi: 10.1177/2053951720939605).
Patton, M. Q. 2009. *Qualitative research & evaluation methods*, Thousand Oaks, Calif.: SAGE.
Power, D. J., Cyphert, D., and Roth, R. M. 2019. "Analytics, bias, and evidence: the quest for rational decision making," *Journal of Decision Systems* (28:2), pp. 120-137 (doi: 10.1080/12460125.2019.1623534).
Provost, F., and Fawcett, T. 2013. "Data Science and its Relationship to Big Data and Data-Driven Decision Making," *Big data* (1:1), pp. 51-59 (doi: 10.1089/big.2013.1508).
Reay, T., Golden-Biddle, K., & Germann, K. 2006. "Legitimizing a New Role: Small Wins and Microprocesses of Change," *The Academy of Management Journal* (49:5).
Salminen, J., Milenković, M., and Jansen, B. J. 2017. "Problems of Data Science in Organizations: An Explorative Qualitative Analysis of Business Professionals' Concerns," in *Proceedings of the 17th International Conference on Electronic Business (ICEB17)*, E. Y. Li and K. Ning Shen (eds.), Dubai. December 4-8.
Sambasivan, N., and Veeraraghavan, R. 2022. "The Deskilling of Domain Expertise in AI Development," in *CHI Conference on Human Factors in Computing Systems,* New Orleans LA USA. 29 04 2022 05 05 2022, New York, NY, USA: ACM, pp. 1-14 (doi: 10.1145/3491102.3517578).

Sarker, S., Xiao, X., Beaulieu, T., and Lee, A. S. 2018. “Learning from First-Generation Qualitative Approaches in the IS Discipline: An Evolutionary View and Some Implications for Authors and Evaluators (PART 1/2),” *Journal of the Association for Information Systems* (19), pp. 752-774 (doi: 10.17705/1jais.00508).

Sheppard, S., Colby, A., Macatangay, K., and Sullivan, W. 2006. “What is engineering practice?” *International Journal of Engineering Education* (22:3).

Smith, S., Garavan, T. N., Munro, A., Ramsey, E., Smith, C. F., and Varey, A. 2021. “An exploration of the professional and leader identity of IT professionals transitioning to a permanent hybrid role: a longitudinal investigation,” *Information Technology & People* (34:2), pp. 789-811 (doi: 10.1108/ITP-02-2019-0084).

Soh, C., Kien Sia, S., Fong Boh, W., and Tang, M. 2003. “Misalignments in ERP Implementation: A Dialectic Perspective,” *International Journal of Human-Computer Interaction* (16:1), pp. 81-100 (doi: 10.1207/S15327590IJHC1601_6).

Strauss, A. L. 1978. *Qualitative Analysis for Social Scientists*, Cambridge University Press.

Strich, F., Mayer, A.-S., and Fiedler, M. 2021. “What Do I Do in a World of Artificial Intelligence? Investigating the Impact of Substitutive Decision-Making AI Systems on Employees' Professional Role Identity,” *Journal of the Association for Information Systems* (22:2), pp. 304-324 (doi: 10.17705/1jais.00663).

Sykes, T. A., Venkatesh, V., and Johnson, J. L. 2014. “Enterprise System Implementation and Employee Job Performance: Understanding the Role of Advice Networks,” *MIS Quarterly* (38:1), pp. 51-72 (doi: 10.25300/MISQ/2014/38.1.03).

Tams, S. 2022. “Helping Older Workers Realize Their Full Organizational Potential: A Moderated Mediation Model of Age and IT-Enabled Task Performance,” *MIS Quarterly* (46:1).

Vaast, E., and Walsham, G. 2005. “Representations and Actions: The Transformation of Work Practices with IT use,” *Information and Organization* (15:1), pp. 65-89 (doi: 10.1016/j.infoandorg.2004.10.001).

van de Ven, A. H., and Huber, G. P. 1990. “Longitudinal Field Research Methods for Studying Processes of Organizational Change,” *Organization Science* (1:3), pp. 213-219 (doi: 10.1287/orsc.1.3.213).

van den Broek, E., Sergeeva, A., and Huysman Vrije, M. 2021. “When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring,” *MIS Quarterly* (45:3), pp. 1557-1580 (doi: 10.25300/MISQ/2021/16559).

van Giffen, B., and Ludwig, H. 2023. “How Siemens Democratized Artificial Intelligence,” *MIS Quarterly Executive* (22:1).

Waardenburg, L., Huysman, M., and Sergeeva, A. V. 2022. “In the Land of the Blind, the One-Eyed Man Is King: Knowledge Brokerage in the Age of Learning Algorithms,” *Organization Science* (33:1), pp. 59-82 (doi: 10.1287/orsc.2021.1544).

Walsham, G. 1995. “Interpretive case studies in IS research: nature and method,” *European Journal of Information Systems* (4:2), pp. 74-81 (doi: 10.1057/ejis.1995.9).

Willcocks, L. 2020. “Robo-Apocalypse Cancelled? Reframing the Automation and Future of Work Debate,” *Journal of Information Technology* (35:4), pp. 286-302 (doi: 10.1177/0268396220925830).

Yin, R. K. 2010. *Case Study Research*: *Design and Methods*, Los Angeles, California: SAGE.

Zhang, Z., Nandhakumar, J., Hummel, J. T., and Waardenburg, L. 2020. "Addressing the Key Challenges of Developing Machine Learning AI Systems for Knowledge-Intensive Work," *MIS Quarterly Executive*, pp. 221-238 (doi: 10.17705/2msqe.00035).