



# Design Intelligence - Pitfalls and Challenges When Designing AI Algorithms in B2B Factory Automation

Jennifer Heier<sup>1</sup>✉, Jan Willmann<sup>1</sup>, and Karsten Wendland<sup>2</sup>

<sup>1</sup> Bauhaus Universität Weimar, 99423 Weimar, Germany  
{jennifer.heier, jan.willmann}@uni-weimar.de

<sup>2</sup> Hochschule Aalen, 73430 Aalen, Germany  
karsten.wendland@hs-aalen.de

**Abstract.** This paper presents results from the on-going research in industrial AI, namely a use case from Business-to-Business (B2B) factory automation, focusing on Artificial Intelligence (AI) technology to predict specific time series for factory planning. This research wants to shed light on the progress in industrial AI applications and, most importantly, the respective design and development process. Assuming that a lack of human-centered-design is a key source for considerable pains of the users, e.g. to trust in the system, and problems for stakeholders involved, e.g. exaggerated expectations [1, 2]. The paper outlines a) a B2B-factory planning case study, including setup and implementation, b) the validation of the process and its results through interviews, and c) the pitfalls and challenges of designing AI algorithms in the given context. Of particular importance are 14 themes that have been developed from analysis of a qualitative study with the development team, users and stakeholders involved. Some of the themes are relevant to other use cases and domains as well, some are use case specific and some derive from the focus of the design perspective. Those insights serve as a basis for further steps of investigation and, ultimately, foster a (cross-disciplinary) transfer of knowledge towards other domains.

**Keywords:** Machine learning · UX · Factory automation

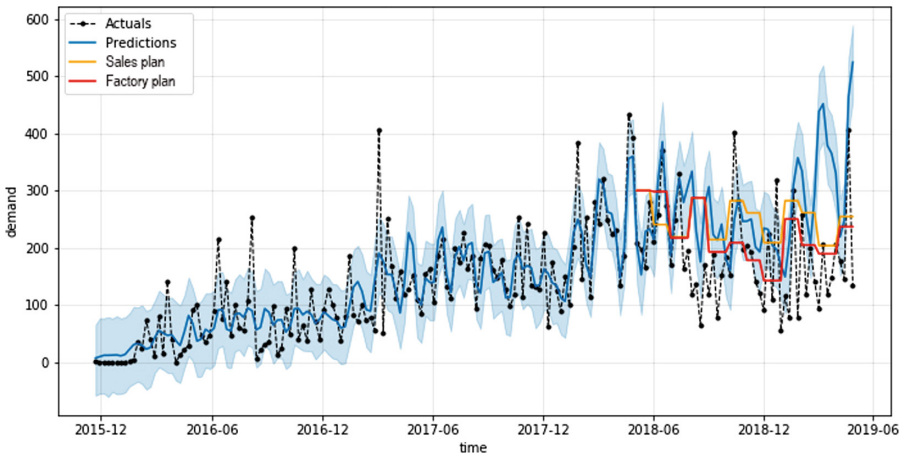
## 1 Introduction

Due to the unprecedented advancement in technology and availability of Big Data, Artificial Intelligence (AI) has become a major factor in business development and optimization, including for example predictive maintenance for machinery, or AI-driven marketing campaigns with customized content for the user [3, 4]. Whereas the use of AI in commercial applications is constantly progressing and research is becoming widely accessible and documented, the world of industrial AI applications and AI agents is a white spot [5]. In the commercial domain, the creative community - such as User Experience (UX) design - already discusses and develops novel B2C applications, and thus pointing to opportunities, as well as challenges of design in the age of AI [6–10]. In fact, design innovation and diversity are considerably lacking in

the domain of data science and machine learning community [11, 12], either in B2C, even more in the industrial sector. This paper therefore describes key paradigms of designing with AI and analyzes a specific use case in the domain of B2B factory automation. It provides an outlook to further points of investigation, as well as a starting point for solutions to the given challenges.

## 2 Context

The portfolio of the respective company focuses on the B2B industrial automation market. Customer services, software, logistics, and hardware are the focus areas. The described use case is based in the hardware unit and its goal is to improve and optimize the factory planning process of a production site for industrial controls in Germany [13]. The current factory planning process is done manually by a group of human planners. Each planner is responsible for a certain number of products, 1700 in total. Information and data from sales, material procurement, capacity of the factory and customer delivery wish dates need to be taken into consideration and influence the process. Their plan is conducted 52 weeks into the future, in order to plan how many pieces of each product need to be produced in order to keep the customer wish date. For the advancement of production, time series prediction with neural networks has been chosen and presents the given technology for the use case [14]. In this, a data model is trained with historic data (consisting of actual pieces sold in the past), tested and validated. If a model has a certain output quality according to the training and testing data it is used to predict 52 weeks into the future, resulting in so called ‘predictive demand planning’ (Fig. 1).



**Fig. 1.** Visual plot of the training (2015-12–2018-06) and validation (2018-06–2019-05) of a neural network for one specific product; x-axis is time and the y-axis are the demand in pieces.

In order to understand and analyze the development process of the described use case, qualitative research methods were used. The qualitative approach is necessary not only to understand why circumstances have been perceived in a certain way, but also to detect unknown topics. In addition, human-centered-design research follows the principle of a divergent way of thinking [15]. As such, 8 one to one (semi-) structured interviews were conducted. Their duration was ranging from 60 min up to 90 min, mainly in person with the respective project team members. The different roles of the interviewees ranged from development team, users and other stakeholders involved. The structure of the interviews was divided in three main parts. First, a section with general questions about the overall target of the project, the role of each interviewee, as well as the AI expertise, second a process- based part with a focus on each step of the development and implementation process, including related issues and challenges, as well as room for future improvement, and third a section on lessons learnt and the underlying role of design in the overall process. Additional knowledge about the design perspective is drawn by the fact that the researcher was a part of the development team as well. The roles of the different disciplines represented by the team members vary from data scientists (with a heavy background in statistics and Machine Learning), a data analyst, a machine learning engineer, a UX designer, a finance specialist, the business domain experts on the clients' side, as well as the production planners and their managers.

### 3 Results

Analyzing the interviews by going through the transcripts and looking for aspects that were important to a majority of the interviewees revealed some initial topics. By grouping the insights and compare them to the overall issues and challenges found in literature review the interviews resulted in 14 clusters. Those 14 clusters can be grouped by their relation a) to the given use case, b) their relevance for AI projects and development in general, and c) to the design expertise (see Table 1). An interesting aspect found, was the fact that all interviewees described a different process, some more similar to each other than others. However, it became obvious that it is difficult to find a common process model for all project team members involved.

The order of the following topics derives from the importance for the different people interviewed. Starting with those which were worth mentioning from all interviewees; development team (DT), users (U), as well as other stakeholders (S), to the ones that were more important to a smaller set of the participants. This is also indicated in detail by the letters associated to the cluster, as well as a choice of quotes taken from the interview transcripts.

#### 01 AI-expertise (DT, U, S)

*“Compared to the beginning of the project, I gained a lot of knowledge about the technology.”*

**Quote 1.** A participant related to the user group

The participants were asked to rate their own AI expertise on a scale from 0 (low) to 10 (high). The reference was up to themselves. Interestingly, nobody rated her-/himself

a 0 or 10. Whereas the data scientists compared themselves towards experts in the field of AI or machine learning and therefore scored around 7 or 8, the business domain experts surprisingly scored themselves alike. Due to the fact, that they compared their expertise from the beginning of the project to the steep learning curve they made towards the point of the interview. The data scientists agreed that a certain level of knowledge is helpful and a key requirement to realize a successful project in the AI-B2B context (including, for example, used technical terminologies, the expertise to rate and evaluate the quality of the AI algorithm, expectations and prioritization of specific agent functionalities).

## 02 Iterative working mode (DT, U, S)

*“To keep the sprints and present results on a regular basis was key for the success of this project.”*

**Quote 2.** A participant related to the development team

Due to the complexity and scale of such an industrial AI project, an iterative approach to develop and implement an AI algorithm is necessary. In the interviews, all participants agreed on this factor. However, while at this stage it is not of key importance which specific iterative method is used (e.g. SCRUM), all involved members and stakeholders agree on the iterative approach. Regular meetings to present interim results and discuss next steps are crucial for the process. And again, if the iterative working method is known to all team members, the more efficient the project.

## 03 Feedback structure, structured feedback (DT, U, S)

A very important aspect around the success of implementing an AI algorithm is the possibility to give feedback and to decide whether to use the data from the AI forecast or not. In the given use case, the business domain expert asked the planners and users for their direct feedback. Since this feedback had no specific structure or format, it was difficult to be considered by the technical development team. Also, an automated approach to include the feedback into the AI forecast was missing and therefore feedback from the planners and users meant extra effort with no significant impact on the AI system and its further development. Quite logically, this caused frustration and a lack of motivation to provide feedback on the planners and user side.

## 04 Definition of Design (DT, U, S)

*“To me design is the look and feel of a product ... But I know this is not what you as a UX person do.”*

**Quote 3.** A participant related to the development team

One question in the interview was targeted towards the interpretation and definition of design. It became quite clear, that in the given cultural setting (Germany), design is perceived as making things look nice and aesthetic. This might be different in the English-speaking world. Rephrasing the question was necessary. Human-Centered-Design and UX as a concept served the purpose better. It evokes the idea of the human as a focus for any step in the process. Need-driven development, instead of primarily technology-driven aspects were perceptions in that context. This clarification and course correction served as a basis for further investigations.

### 05 Visualization (DT, U, S)

Another aspect that plays an important role is visualizing data and output of the AI system. In the given use case line charts and different dashboards in tableau [16] were used. Primarily to communicate the results among all stakeholders involved. Internally the tech team (data scientist, engineers and analyst) also used it to understand the raw data and to evaluate the models. The tool was not used by the planners and users, due to the fact, that the graphics are not helpful for the planning process and the manufacturing site needs numbers. It was an additional tool in their process they refused to use. The value add of a visual representation of the results was not clearly communicated or understood by all stakeholders involved.

### 06 Planning process (DT, U, S)

Process analysis of the current manual planning was conducted. It became quite clear that the current process is one source for the low-quality planning outcome, because it is influenced by subjective human behavior and bias. The development team was lacking the expertise to also address this issue. On top it was not perceived as an important aspect or step by the management. So, the process was left as given and the AI forecast had to fit into the given process. The result is an extra line with a figure the users and planners must evaluate and take into consideration for the factory plan. It is perceived as an extra effort, not as an optimization of the workflow.

### 07 Culture and mindset (DT, U, S)

*“It is really challenging to implement a new technology into old structures. The users need space and time to adapt to the new way of working.”*

#### Quote 4. A participant related to the development team

The exchange with the planners and users was very hard. Due to different aspect of the project set up, but also due to the fact that the AI solution had to fit in the given company structures and work culture. On the long run, AI implementations will perform best, when not only a solution is produced and added to the given infrastructure, but also the structures need to change. Implementing AI needs a change in company culture and people mindset.

### 08 Expectations (U, S)

*“I expect 90% accuracy by the AI prediction, compared to the manual factory plan.”*

#### Quote 5. A participant related to the stakeholder group

When it comes to the evaluation of the AI output (time series prediction), the business domain experts expect a very high accuracy regarding the predictions of the AI algorithm, clearly outperforming the human planner. From a statistics perspective, however, this expectation is unrealistic. The models are trained on historic data. This data is deliberately cleaned from so called ‘outliers’ (meaning data that is unusual, like a big amount of orders that one client made due to some unknown reasons is eliminated from the data set). In turn, if those outliers would be integrated in the predictive model as a pattern, wrong forecasts would be generated, and as a consequence, such data sets are smoothed out in practice. Nevertheless, research showed that the business domain experts expect the AI algorithm to detect those special orders in advance. This is

impossible. Especially time series predictions are very hard to evaluate in advance, therefore a realistic expectation management is a very important aspect. Some products with a high number in orders are easier to predict than products with a small volume or even customized configurations. AI forecasts serve as a value add, especially because it is based on data and not on explicit or implicit human intervention but will not independently decide which number of products to plan.

### **09 Starting point (U, S)**

The project was initiated by upper management which wanted to improve the outcome of the factory planning process. Without directly involving the planners a first technical study was created as a Proof-of-Concept (PoC). However, since planners and users were not integrated from the beginning the PoC had not comprehensive value and its significance has been fairly limited. This is not an AI specific challenge and represents a well-known problem for any project set up. In the described use case, however, it added additional problems, for example, when it came to trust into the output of the AI algorithm and its representative validity.

### **10 Trust in the output (U, S)**

Although expectations for the functionality of the developed AI algorithm are high, it doesn't necessarily mean that humans trust the outcome. For instance, fear of losing job and power are involved, as well as traceability of one's own activities and the unknown mechanisms behind the technology. A testing phase, where manual figures and the AI forecast were used in parallel, helped to ensure validity and trust into the overall approach in the given use case. Unfortunately, there was no structure in place to explain the AI algorithm to the planners and users, which would have supported the acceptance in addition.

### **11 Biased presentations (DT, S)**

*"In hindsight, I think we preferred to show the line charts of the products where the AI predictions performed really well."*

**Quote 6.** A participant related to the development team

The data scientist perspective revealed another very important aspect. When asked in the interviews for the means and frequency of the regular interim presentations, data scientists realized that they showcased the positive outcomes over the negative ones, which led to considerably biased presentations. They justified this 'mechanism' by the fact that business puts pressure on the outcome and hence 'data victory' is rated over failure in such AI projects. At the same time, they realized that this behavior strikes back, for example, when it comes to meeting and managing the expectations of the business domain experts. As provided by the interviews, there seems to be a fine line between implementing AI into such (industrial) contexts and the choice of the right use case combined with the data interpretation which justifies this means the most.

### **12 Gap between prototype and implementation (DT, S)**

*"We had to put a lot of effort and time in the migration of the PoC into a stable productive system."*

**Quote 7.** A participant related to the development team

In the given use case, the first PoC was a selected set of initial 25 products (out of 1700) to test the time series prediction performance. The selection was made by the business domain expert and aligned with the data scientist. As described above, data was cleaned and a first set of models was trained, tested and evaluated. The outcome was surprisingly accurate and thus had been decided to develop and to use the technology for the overall factory automation planning process. In the course of the project, it became clear that the given infrastructure (cloud environment, data security) of the previous PoC was not sufficient and thus a new infrastructure was set up. The new setting didn't perform as well, as the initial experimental environment. A lot of time and capacity was allocated to create a productive setting within the given constraints of data security and company structures. The interviews revealed that this was a pain for all team members involved and extended the final implementation of the project.

**13 Orient, manage, prioritize, eliminate (DT, S)**

UX in the AI project is seen as a way to guide activities in a way that the outcome is the best solution for all stakeholders involved. Although it is clear that not everyone will be satisfied it can also be used as a vehicle to communicate certain decisions, based on needs and not on personal preferences. It was very obvious from the interviews that UX should also help to focus on the most important features, meaning that it should also eliminate wishes and features that are not necessarily needed.

**14 UX and timing (DT, S)**

*“UX really is about the right timing... if it comes too late in the process it cannot influence the direction anymore.”*

**Quote 8.** A participant related to the stakeholder group

The timing for any design or user research activities is very important. All team members agreed on the importance to do the user research at the very beginning of the project, or latest after the successful first technical feasibility study (PoC, Prototype). Only at the end of the project this value add was seen. In the case of early adoption, it could help to support prioritizing the backlog, write user stories and understand the overall process better. It is also clear, that due to the complexity and unforeseen challenges along the way, research is an ongoing process and cannot evaluate all features and needs of the planners and users firsthand. It is perceived as a negative aspect when introduced late in the process, when the team had decided on a way to go already.

**Table 1.** Overview of the 14 themes and their classification into their relevance to the a) use case, b) AI overall, c) design.

Relevance	Clusters
Use case specific	05, 06, 09, 11, (12)
Overall AI relevant	01, 02, 03, 07, 08, (09), 10, 12
Design specific	04, 13, 14

## 4 Discussion of Results

The results above clearly show that there are significant challenges when it comes to the development of AI algorithms in the context of industrial applications. Some of them are not necessarily design-specific and/or B2B-use case specific but need first of all a) to be acknowledged b) to be validated and c) to be addressed in further steps of development and future projects.

The qualitative approach was very helpful to discover the range of different challenges. It opened up the perspective of purely design specific issues. This is also clearly represented by the 14 themes that show a wide spread of relevant aspects to take into consideration while designing AI algorithms. Since the data sample was limited to the given use case and respective development team, it needs to be further elaborated if the 14 themes are sufficient or will be enriched in number and size with further research. Some of the found issues were expected outcomes of the interviews, since the literature review already revealed some overall relevant AI specific challenges. The insight, that there seemed to be not one single process set-up that was valid for all interviewed participants, was unexpected. Furthermore, this shows how important the study of the overall development and implementation process of AI algorithms in the industrial domain really is.

## 5 Conclusion

The insights derived from the research show that some of the challenges are also relevant in other projects, meaning not use case specific [3]. It seems that some issues around AI have an overall validity not based on domain or use case. Expectation management is often an issue when it comes to the impact of AI for example [1, 2] (others see Table 1).

Therefore, this research contributes to the scientific discourse and community by demonstrating that some of the already discussed issues around AI are also valid in industrial AI applications and contexts. That implies, possibly adopting models and techniques in order to address the mentioned challenges, as well as transferring them from this research to other domains.

Other themes are more specific for the given use case, which can be due to the fact, that the domain is industrial AI and not much research is published so far. It also shows that developed design principles or any form of design solution needs to be implemented in a specific context. Here, forms of conceptualizing or generic abstraction miss the specific problem (while being difficult to transfer to other domains and use cases). Thus, the State-of-the-Art has been extended through the focus of this paper into the given domain. Starting to provide insight into the processes and current status of the industrial AI landscape.

The research and conducted interviews show also that some topics are design specific and it needs to be clarified how designers can add value to this process. An evaluation on which ones are the most design specific tasks can be a next step of research. Additionally, investigations into other very similar use cases will be pursued,



elucidating how other teams have fostered solutions. This research will address the question how others solved those challenges or even expand the problem definition.

Based on the 14 themes some initial ideas for improvements have been derived. Taking the AI-expertise as an example, the project team came up with a concept for a personal training. The value add of such an activity would be the attempt to start the project with a certain level of AI-expertise from all team members involved. Also addressing other issues, like the expectation management for example, as well as the possibility to give every team member a certain role and task to be represented in the training, making the training a cross-disciplinary effort and unified approach. Whereas the overall concept of the training could easily be transferred to other domains and target groups, the specific content of the trainings needs to be very use case specific. This means, in turn, that the data used for presenting certain characteristics and features needs to be very similar to the ones the training participants are used to. In the case of factory planning and demand forecasting, historic data from actual orders are needed. Additionally, talking about time series predictions, doesn't necessarily mean to introduce all machine learning approaches, such as image or voice recognition, because it is not used and necessary in the given context. Further exploration on possible solutions for the development of AI applications in industrial AI needs to be done.

To sum up, understanding and analyzing the initial pitfalls and key challenges of the development of AI agents for predictive demand planning in factory automation offers a strategic and productive starting point for the development and, at the same time, to bring forward a robust method to improve the set up and implementation of similar use cases.

As a next step, the research from other, but very similar, factory automation site activities will be analyzed. Based on the most relevant aspects further research with other, also external experts, will be conducted. Parallel to this, a comprehensive investigation into already given principles and tools from other domains will help to address relevant design and development steps towards future solutions in the B2B factory automation context. This dialogue is so important, because AI algorithms enter every aspect of human life and with this comes great responsibility and power that shouldn't be left to a limited group of people in society.

## References

1. Amershi, S., et al.: Guidelines-for-human-AI-interaction. In: CHI 2019, Glasgow, Scotland, UK (2019)
2. Google. People AI Guidebook. <https://pair.withgoogle.com>. Accessed 24 Feb 2020
3. Chui, M., et al.: McKinsey Global Institute: Notes from the AI Frontier, discussion paper (2018). [https://www.mckinsey.com/~ /media/mckinsey/featured%20insights/artificial%20intelligence/notes%20from%20the%20ai%20frontier%20applications%20and%20value%20of%20deep%20learning/mgi\\_notes-from-ai-frontier\\_discussion-paper.ashx](https://www.mckinsey.com/~ /media/mckinsey/featured%20insights/artificial%20intelligence/notes%20from%20the%20ai%20frontier%20applications%20and%20value%20of%20deep%20learning/mgi_notes-from-ai-frontier_discussion-paper.ashx). Accessed 24 Feb 2020
4. Kureishy, A., et al., Achieving Real Business Outcomes. O'Reilly Media (2019)
5. Dunning, T., et al.: AI and Analytics in Production. O'Reilly Media (2018)

6. Clark, J.: Design in the Era of the Algorithm (2017). <https://bigmedium.com/speaking/design-in-the-era-of-the-algorithm.html>. Accessed 24 Feb 2020
7. Hebron, P.: Machine Learning for Designers. O'Reilly Media (2016). <https://www.oreilly.com/library/view/machine-learning-for/9781491971444/copyright-page01.html>. Accessed 24 Feb 2020
8. Noessel, C.: Designing Agentive Technology: AI That Works for People. Rosenfeld Media, New York (2017)
9. Lovejoy, J., Holbrook, J.: Human-Centered Machine Learning (2017). <https://medium.com/google-design/human-centered-machine-learning-a770d10562cd>. Accessed 24 Feb 2020
10. Taschdjian, Z.: UX design in the age of machine learning (2018). <https://uxdesign.cc/ux-design-in-the-age-of-machine-learning-2fcd8b538d67>. Accessed 24 Feb 2020
11. Li, F.: Put humans at the center of AI. MIT Technol. Rev. **120**(6), 26 (2017)
12. Yang, Q.: The role of design in creating machine-learning-enhanced user experience. In: The AAAI 2017 Spring Symposium on Designing the User Experience of Machine Learning Systems (2017)
13. Siemens AG. Industrial Controls - SIRIUS. <https://new.siemens.com/global/en/products/automation/industrial-controls/sirius.html>. Accessed 24 Feb 2020
14. Nielsen, A.: Practical Time Series Analysis: Predictions with Statistics and Machine Learning. O'Reilly Media, Sebastopol (2019)
15. Design Council. Framework for Innovation (2019). <https://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond>. Accessed 24 Feb 2020
16. Tableau Software. <https://www.tableau.com>. Accessed 24 Feb 2020