

Article

On the Track to Application Architectures in Public Transport Service Companies

Stephan Jüngling ^{1,*}, Ilir Fetai ², André Rogger ^{2,*}, David Morandi ³ and Martin Peraic ¹

¹ School of Business, University of Applied Sciences and Northwestern Switzerland (FHNW), 4052 Basel, Switzerland; martin.peraic@students.fhnw.ch

² Swiss Federal Railways (SBB), Competence Center on Machine Perception, 3014 Bern, Switzerland; ilir.fetai@sbb.ch

³ Regional Transport Bern-Solothurn (RBS), 3048 Worblaufen, Switzerland; david.morandi@rbs.ch

* Correspondence: stephan.juengling@fhnw.ch (S.J.); andre.rogger@sbb.ch (A.R.)

Abstract: There are quite some machine learning (ML) models, frameworks, AI-based services or products from different IT solution providers available, which can be used as building blocks to embed and use in IT solution architectures of companies. However, the path from initial prototypical proof of concept solutions until the deployment of proven systems into the operational environment remains a major challenge. The potential of AI-based software components using ML or knowledge engineering (KE) is huge and the majority of small to medium enterprises are still unsure whether their internal developer teams should be extended by additional ML or KE skills to enrich their IT solution architectures with novel AI-based components where appropriate. How can enterprises manage the change and visualize the current state and foreseeable road-map? In the current paper, we propose an AI system landscape for the public transport sector, which is based on existing AI-domains and AI-categories defined by different technical reports of the European Commission. We collect use-cases from three different enterprises in the transportation sector and visualize them on the proposed domain specific AI-landscape. We provide some insights into different maturity levels of different AI-based components and how the different ML and KE based components can be embedded into an AI-based software development life-cycle (SDLC). We visualize, how the AI-based IT-solution architecture evolved over the last decades with respect to coupling and decoupling of layers and tiers in the overall Enterprise Architecture.

Keywords: AI architecture; AI maturity model; AI landscape; machine learning; knowledge engineering; software development life-cycle; domain-specific AI; AI in transportation industry; enterprise architecture; MLOps



Citation: Jüngling, S.; Fetai, I.; Rogger, A.; Morandi, D.; Peraic, M. On the Track to Application Architectures in Public Transport Service Companies. *Appl. Sci.* **2022**, *12*, 6073. <https://doi.org/10.3390/app12126073>

Academic Editor: Antonella Petrillo

Received: 11 March 2022

Accepted: 12 May 2022

Published: 15 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

When looking at the evolution of software architectures that started from mainframes and dumb 3270 terminals to N-Tier architectures, the distribution of specialized functionality to dedicated hardware-related tiers such as database servers or logical layers such as the presentation layer, which was triggered by a substantial technology gap between HTML as a markup language for the graphical user interface (GUI) and the object-oriented programming languages used to implement the application- and business-logic. With an increasing focus of organizations to business processes, workflow and rule engines were incorporated into the software architectures, leading to additional layers or tiers. With the increasing capabilities of artificial intelligence (AI), which in itself can be decomposed into machine learning (ML) and Knowledge Engineering (KE) sub-domains, additional layers and tiers slowly establish and need to be incorporated into the overall IT architecture and the software development life-cycle (SDLC) of all these different building blocks. In case of ML models, an additional challenge is added to the life-cycle management of components, which is equally severe as the gap between the markup and the programming

languages in the past. The models no longer can be designed, developed, tested and put into production when being compliant to their specification, but their quality is rather validated based on probability scores and, even more challenging, can modify its behavior or learn during its operation and potentially degrade over time due to data and concept shifts [1]. Such scenarios are new and require not only an additional focus to the initial maturity of the components but are also challenging the maturity of an organization with respect to managing these components over its life-cycle. How can an enterprise manage these issues? What are appropriate changes in the software development life-cycle when AI-based components, that are trained instead of programmed need to be incorporated in the IT-Architecture? How can the maturity of the components be tracked during an experiment-driven and data-driven development process? How can it be monitored in the runtime environment? We will analyse and visualize these changes from a structural as well as a behavioural perspective. Maturity models have been applied in many different domains. It is quite common to specify the maturity of organizations with respect to business process management, whether the processes are explicitly modeled with BPMN [2], implemented with the help of workflow engines. It is even possible to extract dedicated business rules from the processes. The rules can be modelled with the decision model and notation DMN [3] and handed over back to the responsibility of the business units, where they can be quickly modified according to their needs. Maturity models such as the CMM or BPMM have been developed in order to characterize the different awareness level, to which software development processes or business processes are managed in enterprises [4]. More recently, maturity models seem to gain focus for the assessment of AI-based software components [5] such as the one of Gartner [6] or Forrester [7]. Based on general AI-categories and domain specific railway related sub-themes, we first focus on creating an overview of AI-based components in public transport service industries from a structural perspective, resulting in a domain-specific AI-landscape. Second, we visualize, how AI-based functionalities are embedded in appropriate layers and tiers in the context of a domain-independent evolution of the overall IT Architecture. Third, we provide domain-specific examples and different use-cases of AI-based components, how they are developed and embedded into the organization's software development life-cycle from a behavioral perspective, which is driven by the different maturity levels from initial prototypical implementations until the operation of AI-based components in production. Other than previous papers, that do an in-depth analysis of current challenges of incorporating AI-capabilities into the software engineering process from Microsoft in large scale [8], smaller organizations, which work in a specific business domain such as the transportation service sector, face similar challenges. However, they are most probably starting from a lower maturity level, both from the perspective of the organization, as well as from the perspective of the maturity level of their AI-based software components. Our paper collects cases from three different enterprises, in order to develop a common understanding of the necessary changes to implement AI-based application architectures in a specific business domain such as the public service transportation sector. For this purpose, we start with some general definitions based on the technical reports of the European Commission before we dive into some theoretical aspects of modelling the structural as well as behavioral aspects of AI-related software engineering. Based on these insights, we present different use cases of the different organizations, which all can be placed on a proposed AI landscape for the public transportation business domain.

2. Current Frameworks and Maturity Models Which Could Be Used to Build Up an AI Landscape in Organizations

The recent advancements in AI research in general and machine learning in particular have triggered broad interest also to global economies, to entire business domains and to governments and politicians. In the global context of Industry 4.0, AI is part of the EUs Digital Single Market Strategy and in the public transport sector, AI is a main topic in the UIC (International Union of Railways). However, as stated by a recent technology report

of the EU, there is up to now no single, commonly agreed definition of AI but a rather huge variety of different definitions that emerged from early 1955 until today [9]. This report is focusing on an operational definition of AI where they are listing different AI core domains (CD) representing fundamental goals of AI as well as transversal domains (TD) referring to common issues of integrating AI core domains across disciplines and different knowledge areas. Both, core and transversal domains are listed in Table 1 with their appropriate sub-domains.

Table 1. AI Core and Sub-Domains.

Domains	Sub-Domains
CD—Reasoning	knowledge representation, automated reasoning, common sense reasoning
CD—Planning	planning and scheduling, searching, optimization
CD—Learning	machine learning
CD—Communication	natural language processing
CD—Perception	computer vision, audio processing
TD—Integration and Interaction	multi-agent systems, robotics and automation, connected and automated vehicles
TD—Services	AI Services
TD—Ethics and Philosophy	AI Ethics, Philosophy of AI

Based on these domains and subdomains, they further collected different keywords which are relevant for the domains (e.g., case based reasoning, causal inference, expert systems etc. to name a few in the domain of reasoning) which can be used to build an entire AI taxonomy that is transparently mapped to the extensive list of definitions of AI between 1955 and 2019. In a second report from AI Watch, they list the AI-Categories with corresponding technologies as shown in Table 2. Furthermore, they define nine different Technology Readiness Levels (TRL, Table 3) with their appropriate descriptions [10].

But in addition to the technology readiness, organizational readiness factors need to be considered in the process of AI adoption, such as those considered by Jölnk et al. [11], who compiled five categories with 18 different AI readiness factors (See Table 4) out of exploratory expert interviews as well as findings from scientific and practitioners literature.

Aside of the conceptualization framework, they point out the importance that technology readiness and organizational readiness are “highly interdependent and mutually reinforcing”. But how should this problem be solved by organizations? Which would be the most adequate technologies to start with initial prototypes? Should each organization come up with their own AI map or AI landscape in order to visualize the current state of usage of AI components? In the following, we will show, how a domain specific template can be compiled, which could be reused by individual companies, and help them to spot out unused AI potentials.

Table 2. AI-Categories with corresponding technologies.

Categories	Technologies
Knowledge Representation & reasoning	Expert Systems
Learning	Recommender Systems, apprentices by demonstration
Communication	Machine Translation, Speech Recognition
Perception	Facial recognition, text recognition
Planning	Transport and scheduling systems
Physical Interaction (Robotics)	Self-Driving Cars, Home Cleaning Robots
Social & Collaborative Intelligence	Negotiation Agents
Integration Technology	Virtual Assistants

Table 3. Technology Readiness Levels (TLRL).

Technology Readiness Levels	Descriptions
TRL 1—Proof of concept	Basic principles observed
TRL 2—Proof of concept	Technology concept formulated
TRL 3—Proof of concept	Experimental proof of concept
TRL 4—Prototype	Technology validated in lab
TRL 5—Prototype	Technology validated in relevant environment
TRL 6—Prototype	Technology demonstrated in relevant environment
TRL 7—Prototype	System prototype demonstration in operational environment
TRL 8—Product or Service certified	System complete and qualified
TRL 9—Deployment	Actual System proven in operational environment

Table 4. Organizational AI Readiness Factors.

Categories	AI Readiness Factors
Strategic alignment	AI-business potentials, Customer AI readiness, Top management support, AI process fit, Data-driven decision making
Resources	Financial budget, Personnel, IT infrastructure
Knowledge	AI awareness, Upskilling, AI ethics
Culture	Innovativeness, Collaborative work, Change management
Data	Data availability, Data quality, Data accessibility, Data flow

3. Compiling an AI Landscape for the Public Service Transportation Business Domain

In order to build such an AI landscape for the railway transportation service industry, we propose to reuse the definitions from AIWatch from the previous chapter and combine it with the AI knowledge map (AIKM) of Corea [12]. Corea's 2-dimensional AIKM uses the AI-domains on the vertical and the AI-paradigms (e.g. symbolic, statistical, subsymbolic) on the horizontal axes. This AIKM has been reused as a pragmatic definition of AI technologies for the debate of AI ethics in healthcare [13] as well as a starting point for a paper addressing future applications of AI based on insights, risks, success stories and failures from the past [14]. In order to generate a domain specific landscape, the very high-level and general purpose dimension of the AI-paradigms can be replaced domain specific sub-themes, which were defined in the context of the Transport Research and Information Monitoring and Information System (TRIMIS) [15]. These sub-themes are developed by the European Commission for the development, testing, adoption and integration of new technologies into the railway industry and published a Joint Research Centre (JSR) report. Although not focusing explicitly on the incorporation of AI-based components, the sub-theme clusters of the EU's railway related research and innovation (R&I) framework seem to be standardized in the transportation domain and can be reused: The original domain independent AIKM places the AI technology solutions on a two dimensional map with AI Problem domains on one axis and AI paradigms such as symbolic AI, statistical AI and sub-symbolic AI on the other. Since we want to compile a domain specific AI landscape, we want to focus on specific AI components rather than the general AI technology solutions. In order to do that, we keep the AI problem domains and replace the AI paradigms by the domain specific problem sub-themes identified by TRIMIS and listed in Table 5.

Table 5. Rail R&I sub-themes.

Sub-Theme	Sub-Theme Focus
Efficient stock design and manufacturing:	Innovations in the area of rail stock (passenger trains, freight locomotives and wagons) in terms of both design and manufacturing.
Emissions and noise reduction:	Research projects investigating the reduction of railway noise and emissions; through reduced energy consumption and reduced railway vibrations causing noise (including track interventions to reduce noise)
Track and stock maintenance:	Technologies to predict maintenance needs of both rolling stock and the railway tracks, as well as techniques for monitoring the condition of each in real-time.
Efficient rail operations:	Improving the efficiency of the railway through traffic management systems, automation and control systems.
Infrastructure management:	Optimal management of railway infrastructure, which includes electrical infrastructure (overhead lines), level crossing safety and protection from potential threats.
Information and Communications Technologies (ICT) solutions to enhance the rail travel experience:	Services to improve the passenger experience, including smart-ticketing systems spanning multiple transport modes and considering passengers with restricted mobility.

In the following sections, we first present the structural aspects of incorporating AI into the IT architectures of transportation service industries, then describe some behavioral aspects, about how AI can be embedded into the organizations in the form of AI-based Software Development Lifecycle and finally visualize how the AI-specific components are incorporated into the technical IT-architecture similar as other heterogeneous tiers or layers in the past.

3.1. AIKM Landscape—The Domain-Specific Structural View

The following AI landscape visualized in Figure 1 shows the domain specific AI landscape for transportation service industries, where we place different AI-components.

As public transport is a very infrastructure-heavy business, it is obvious that the focus of the AI application also lies in the optimisation of its railways infrastructure. Especially in the sub-theme of track and stock maintenance, ML-based components are most suitable. Figure 1 shows only a few concrete examples, some of which, such as AISI (see Section 4.3), are already at the TRL9 operational level, others such as the predictive maintenance of the pantograph-catenary system (Section 4.2), are in early TRL5 prototype levels. AI technologies are used or tested in many other maintenance areas as for example in vegetation and neophyte control or to monitor and ensure safety regulations of passing railway vehicles at camera gantries (Wayside Intelligence).

In the efficient rail operations sub-theme, KE-based components, such as the Traffic Management System can be allocated to the reasoning and planning AI-Domain, while the ML-based Passenger Counting system uses perception and learning to optimize occupancy and the Shunting Communication component is used for more efficient and safer shunting execution by replacing PDF manuals with a voice-controlled information system.

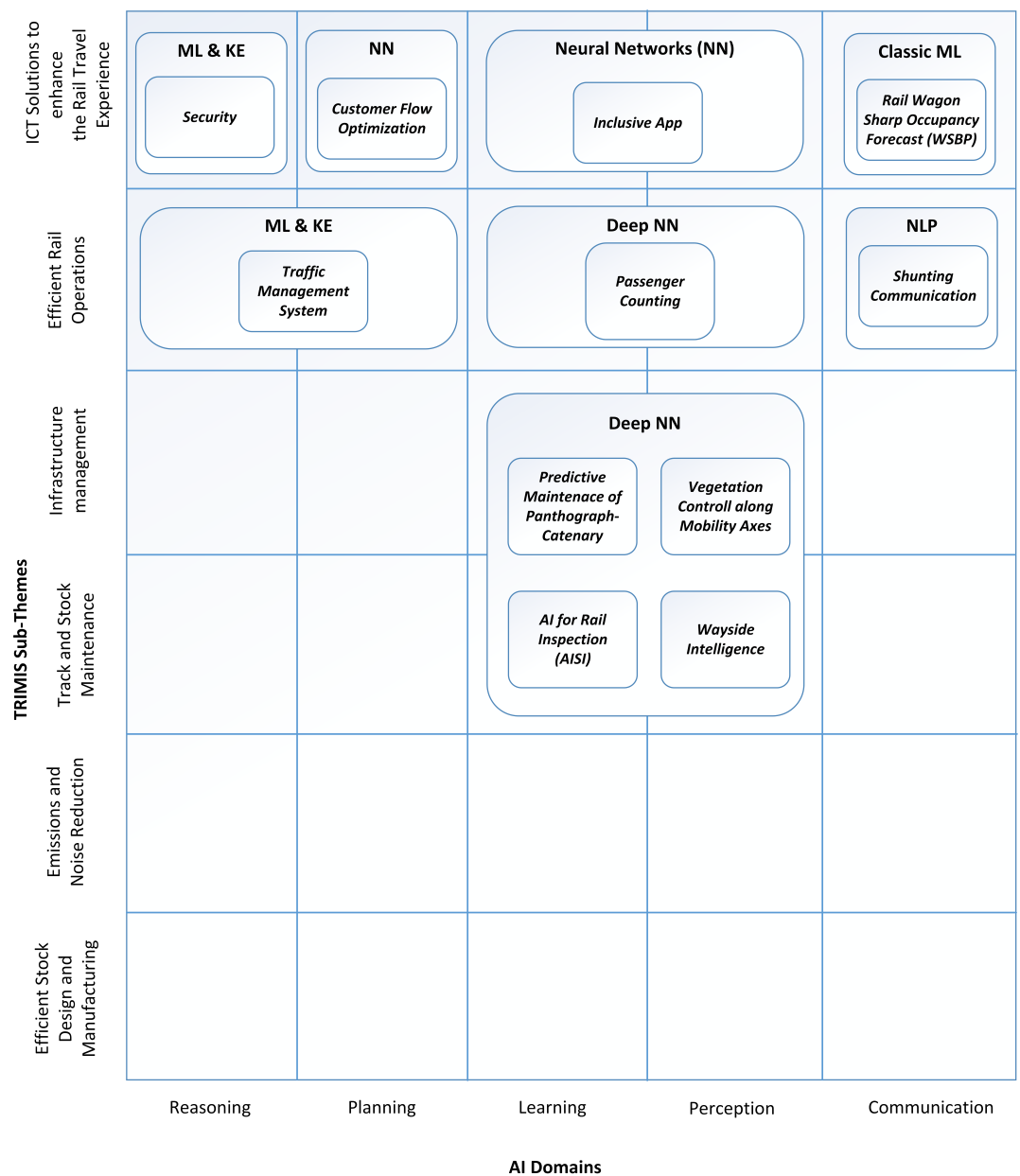


Figure 1. AI landscape in public transportation.

A wide variety of AI applications are conceivable in different sub-themes and some are already in use. Customer Flow Optimization (CFO) can be used in stations to optimise customer flows and increase safety. With the help of an IA, visually impaired people can be guided to the boarding points of their train thanks to haptic signals or with the wagon occupancy forecast WSBP (Wagenscharfe Belegungsprognose), passengers can find out about the expected occupancy in the individual train sectors even before they start their journey.

AI adoption in the public transport can also be seen as strategic alignment in the UIC (International Union of Railways). Their main objectives are security, operations and efficiency in public transport. AI can recognize defect breaks rising the security or can be used for optimizing timetables much beyond the possibilities considered so far. And finally, a lot of assets are maintained today by a fixed time schedule, which means that assets are maintained, even if they are still working well. With AI-enabled predictive maintenance, assets will only be maintained if needed, which saves costs.

Another major application of AI is around improving the customer experience. For example, AI technologies are used to improve customer flows at railway stations, to optimise support for the disabled, which is done by the Inclusive App or to recommend where to find free seats, which is communicated by a WSBP as mentioned before.

The AI landscape can provide an overview of AI-based applications in the domain specific area and help to spot out areas which are already covered by AI or spot out those with open application potential. More detailed descriptions from the three organizations SBB, RBS and BVB will be illustrated in a subsequent section, but first, we will cover further aspects, of how these examples are embedded in a AI-based software development life-cycle, which also needs to be adapted due to the different nature of AI and in particular of ML-based components.

3.2. AI-Based Software Development Life-Cycle—Organizational Behavior View

The novel experimental characteristics of AI, of ML models in particular, as well as the vast amount of data, which needs to be taken into the overall picture have a considerable impact on how the components are developed and managed within the organisation.

The traditional SDLC needs modifications and new roles and responsibilities are created, which influence how the exploratory ML-models evolve from an initial TRL1 proof of concept state to the final TRL9 deployment state. Traditional software development is well established. When the problem is recognized and analysed by the business, business analysts conduct a well established requirements engineering sub-process and hand it over to the software engineering team for implementation, testing and deployment to the runtime environment for production, such as it is visualized in Figure 2 as a BPMN diagram. In the case of engineering AI-based components, the process is less straight forward. Developer skills are well established but Data Science skills or entire new teams need to be integrated into the enterprise. Business analysts need to decide, whether the problem is best solved with traditional Requirements Engineering and subsequent traditional Software Development and well established Deployment pipelines, or whether it should be handed over to the new path of Data Analytics and subsequent Hybrid-AI Engineering where ML and KE have to be combined most efficiently. Once the problem is allocated either to the Machine Learning or Knowledge Engineering sub-process, the different TRL levels from proof of concept until TRL8 need to be evolved, until the production ready TRL9 artefact can be deployed into production. The Software Engineering team is challenged to integrate traditional DevOps with MLOps, and manage deployment pipelines with components from traditional Software Engineering, ML and KE processes. The monitoring of the software components (SCs) in production are also getting more challenging, since other than traditional SCs that have static defects which might be fixed with appropriate static patches, ML models may slightly degrade due to dynamic changes in data and KE models need to be dynamically adjusted to changing semantics in the business domain models.

The Wagon Sharp Occupancy Forecast (WSBP) is a classic ML component placed in Figure 2 in the communication AI-domain to increase customer experience, which is one of the very first productive ML application of SBB. In fact it is a simple recommender system that shows the expected carrier occupancy to the passengers during the ticket purchase. While analysing the problem, it was clear that a simple rule based procedure would not work and that the recommendations need to consider patterns in the data. Thus a proof of concept was made that confirmed the hypothesis and a ML procedure was trained with occupancy data from over a year. This occupancy indication was then integrated into the ticketing application. New occupancy data is continuously added to the data pool and outdated data has to be discarded. In such a way, changes in behaviour, as it was the case with the Covid-19 pandemic, were taken into account and the quality of forecasts could be adjusted within a short period of time.

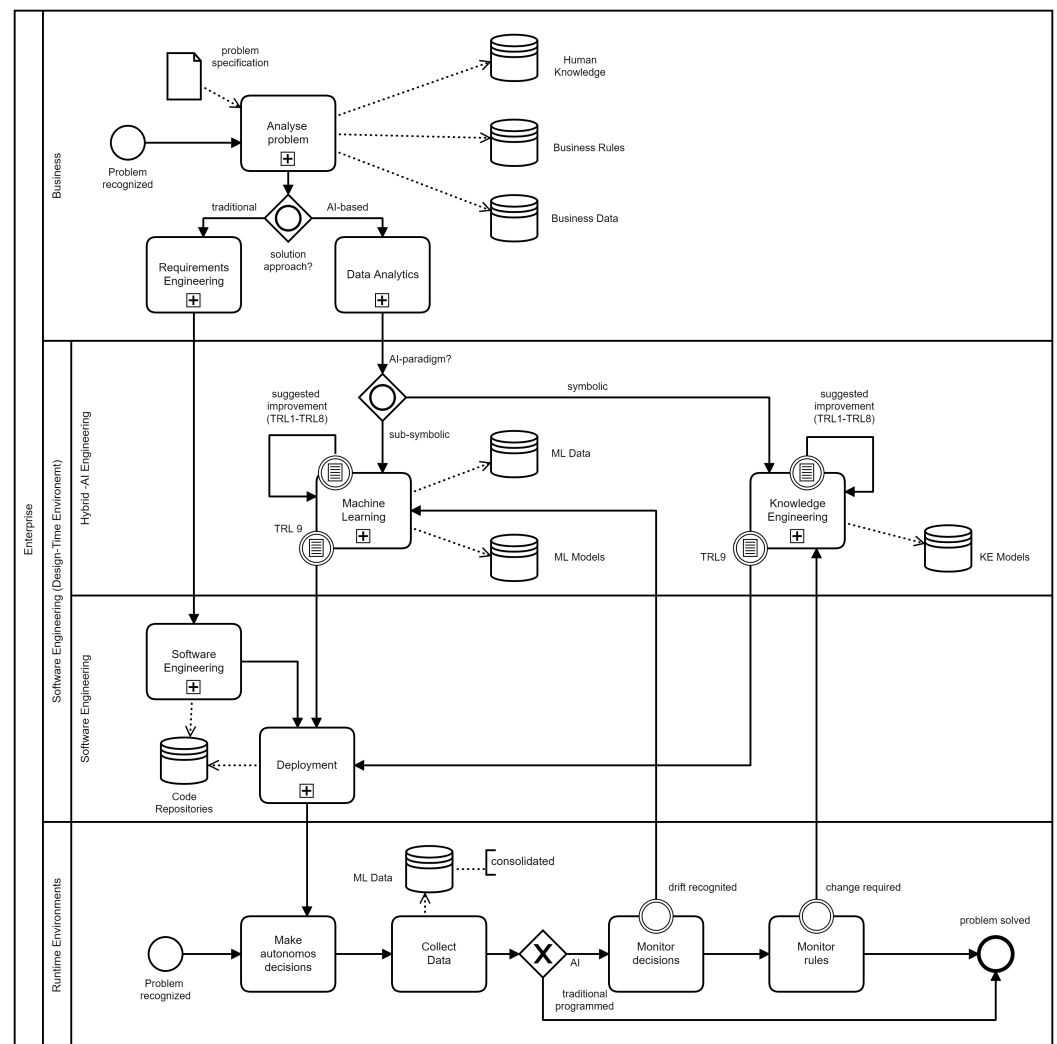


Figure 2. Embedding AI solutions into the SDLC.

3.3. Towards an AI-Embedded Software Architecture

As already described in the introduction section, the incorporation of AI into the IT architecture is about equally important, as the incorporation of dedicated database systems, the global trend of switching to a web-based system architecture or the incorporation of business processes as mature sub-systems, which all need to be treated with care in terms functional coupling and decoupling from the rest of the system components. It needs a major change in the overall architectural design in order to be scalable and manageable over time.

In the context of Industry 4.0, a wide potential of data-driven advancements of existing or new products and services exist and most often also causes challenges not only to adapting the SDLC in the organization, but also to manage the change in the context of the overall Enterprise Architecture. The vision of smart homes, smart cities, smart companies creates pressure to investigate into the application potential of new technologies and poses the challenge for small to medium enterprises (SMEs) to create the necessary awareness to the potential of the new technology trends and continuously monitor the application potential as window of opportunity to further develop their existing business models or create new ones. What Schröder [16] stated for the process of incorporating new technology in SMEs in the production sector can be generalized and as such can also be applied for the public transport service sector. The overall IT Architecture has to absorb the new characteristics of AI in the context of the evolution of layers and tiers, with appropriate possibilities for coupling and de-coupling of the different software components. The more

the IT grows in enterprises, the more important are the aspects of the overall Enterprise Architecture. The AI landscape shown before provides an AI-domain specific view of how the different SCs are related to a domain-specific business domain, in our case the transportation service industries. However, the overall Technical Architecture is domain independent and is also evolving over time, such as visualized in Figure 3. The role of the data in the enterprise, which was the first sub-system that was decoupled as layer or even dedicated tier from the rest of the IT needs to be adjusted caused by both types of AI, ML and KE. In the context of ML, the relevance increased drastically. The quality of the data not only needs to be controlled for the purpose of being aggregated by Business Intelligence (BI) tools from different operational platforms to generate the management reports, which in turn enable to make data-driven decisions and to adjust the organization and the overall business according to customer and market needs, but is meant to be the source of making decisions in the operational businesses. As pointed out in the AI-based SDLC BPMN in the lane of the Runtime Environment, autonomous decisions are made out of AI-based components and potential data- or model-drifts need to be recognized. In the context of security and privacy relevant scenarios, where video data is used for ML-based passenger counting systems, as we will elaborate in the use-cases section, it needs to be insured, that the data only is used on the edge devices for making appropriate decisions and is directly discarded afterwards and never reaches out to any server-based storage, where personal data could cause potential privacy issues. A further most important attention has to be dedicated to the coupling and de-coupling of ML and KE based AI. Although in the past, decoupling of appropriate functionalities, for instance the decoupling of the GUI layer from the DB layer is considered as state-of-the-art quality criteria in traditional software development, the opposite fact seems to hold true for the coupling of KE and ML-based functionality in the AI-domain. The fact that neither ML nor KE alone will lead to most adequate solutions, but that the coupling of both sub-symbolic and symbolic reasoning is essential for a long-term success, is the cause of conducting a conference symposia such as AAAI Make [17], or the reason for issuing a special issue journal such as this one.

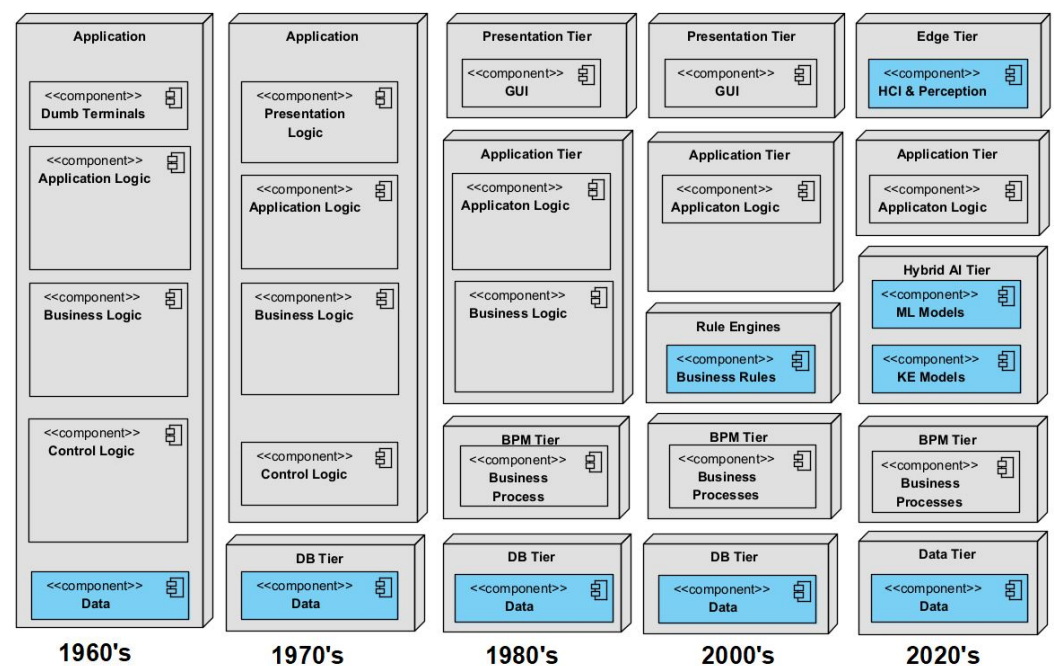


Figure 3. N-Tier Software Architecture incorporating AI based Tiers.

4. Current State of Use-Cases from Different Transportation Service Providers

In the following sections, we will show different stages of ML-based software components from a Software Development Life-Cycle SDLC perspective, starting with two pilot

projects in smaller transportation companies, followed by use-cases, where the AI-based solutions are already deployed in operational businesses of SBB, the largest transportation service provider in Switzerland.

4.1. Use Case—Passenger Counting at BVB

Basler Verkehrs-Betriebe (BVB) operates a network of more than 180 km in the city of Basel. With 85 streetcars and almost 90 buses, nearly 340,000 people are transported daily. This makes BVB the largest partner of the public transport association of northwestern Switzerland (TNW), which provides its customers with barrier free transport to various cantons as well as to the neighboring countries of Germany and France. As a TNW partner, BVB is obliged to record passenger counts in order to determine the percentage of total passengers carried within the TNW network to receive the associated financial compensation. This data is currently collected using light barriers in the infrared spectrum. The optical barriers are interrupted upon passing the entrance and can thus measure and count the number of direct movements into or out of the vehicle. This simple but reliable method is not capable to recognize different object types as bikes or wheelchairs or to detect passenger movements that do not follow the previously described simple movement pattern. Regarding the TNW application purpose mentioned at the beginning, these additional possibilities are not crucial and therefore not desirable for the business at present. An initial analysis of purchasable solutions to solve such complex visual tasks within the CD Learning and Perception was unsuccessful. One of the main reasons behind this fact is that the software supply for the public transport sector is largely monopolized by a handful of software providers. Consequently, emerging technological trends such as AI-based systems are not always immediately available. Thus, the only option left is to develop the systems internally. Yet the small to medium-sized public transport companies are not in a position to achieve a TRL above 7, nor are their internal IT departments. Missing R&D teams and a not fully implemented AI-SDLC lead to that fact. After consultation with the management respectively the FHNW, BVB was able to find a possibility to conduct such a project as a bachelor thesis [18,19]. Thereby a proof of concept of a TRL 6 level should be created to initiate the awareness in the management for AI and its feasibility in the domain of efficient rail operations.

In the run-up to the project, it already became evident that visual sensors used in public spaces in Switzerland face serious obstacles. Due to data protection laws, it is not possible to transfer image data from the public space to a central computer system. Therefore, the first step was to find a way to perform the computationally intensive computer vision tasks in a distributed manner at the edge of the network inside the vehicle. This architecture would allow AI systems to be employed for advanced analytics, despite the legal obstacles. Fortunately, in 2019, various embedded systems optimized for AI operations have been introduced. Furthermore, AI libraries such as TensorFlow or Pytorch were available, which facilitated the development of AI applications. After various discussions with experts from different departments involved in passenger counting at BVB along with use case specific observations, the appropriate setup for the project was defined. The Coral Dev Board produced by Google was chosen as the hardware, together with the suitable AI library TensorFlow Lite. For the subsequent selection process regarding the object recognition model, all available scanners were assessed. The assessment included test cases with the individual models with simulated images from the vehicle. Thereby it soon became evident that both the accuracy of the detection but also the performance of the entire model plays a crucial role when it comes to development on an integrated system. This development approach is consistent in most respects with the AI-SDLC process proposed in Figure 2, except that due to BVB's smaller organizational size, most activities were conducted within a single traditional Software Engineering department, in partnership with the business. Given the limiting hardware capabilities coupled with the ARM architecture of the Coral Dev Board, optimization of any additional components required had to be performed over several iterations, with consideration given to both of these factors. For example, the AI

model for object detection was only executed at a pre-defined frame interval. In between, a less resource-intensive correlation tracker from the DLIB library was utilized. Finally, a simple Centroid Tracker was applied for the re-identification of the objects. This tracker compares the transmitted coordinates and measures on the basis of the Euclidean distance whether it is likely that the object identified in the last frame has made such a movement. Each of the previously mentioned features as well as their partly complex parameters such as the camera placement had to be tested and evaluated in the vehicle as illustrated in Figure 4.



Figure 4. Simulated boarding and deboarding scenarios were performed to measure the accuracy of the newly developed AI as well as the infrared passenger counting system.

Both, common and unique scenarios of passenger movements into and out of the vehicle were studied. This includes individual boarding as well as boarding in larger groups, varying complexity and velocity. With such an agile oriented methodology, a simple but quite effective passenger counter could be developed in a relatively short time. The final comparison between the AI and the infrared system demonstrated promising results even without domain-specific training of the AI-based object recognition model. Nevertheless, it should also be noted that the system in its current design was not fully capable to beat the existing counting system in terms of accuracy.

4.2. Predictive Maintenance of the Pantograph-Catenary System at RBS

Regionalverkehr Bern-Solothurn (RBS) is a Swiss public transport company in the region of Bern and Solothurn. RBS operates a fleet of 50 trains and 48 buses, a train network of 54 kilometers and the corresponding infrastructure. Over 86,000 passengers using RBS' transport services on peak days. RBS is part of the S-train network in the city of Bern and has a very dense and optimized timetable. Therefore, it is important to prevent incidents as efficient as possible. However, faulty infrastructure or rolling stock can cause delays, and even minor disturbances could provoke major disruption to the entire timetable. There are many causes for delays, some of them can be predicted and some of them cannot. Nevertheless, periodic maintenance can help to minimize the risk of delays and accidents. However, they are time-consuming and cause idle times. Furthermore, it is important to know what the source and root cause of the defects are. This allows to determine appropriate maintenance activities.

Today, with the help of sensors, it is possible to collect plenty of data. However, processing and detecting anomalies in the collected data is too complex and time-consuming for a human to perform. In this use case, machine learning could help to minimize this effort.

In order to localize arc ignitions in the pantograph-catenary system during train operation, which is an anomaly, no automatic detection system can be drawn on today. The time of the occurrence, the location, the acceleration of the train or further context data would be interesting to know in order to determine why arcs occur and where the catenary system should be inspected.

To address these challenges the arcVision System prototype was developed in the context of a master thesis at FHNW in corporation with RBS [20]. The system consists of two different parts. The first part is the arcVision Scanner, which is able to collect the necessary sensor and context data, consisting of a single board computer, a GPS receiver, a camera, a temperature and humidity sensor. The second part is the arcVision Model, a Convolutional Neural Network (CNN) used for arc detection of the recorded pictures from the arcVision Scanner. The goal of the prototype was to achieve a TRL6, where the prototype needs be demonstrated in the operational environment and had to be mounted on the roof of a train in order to collect data. A supervised ML approach was used to train the ArcVision Model, which is a CNN based on Keras TensorFlow.

As visualized in the Hybrid-AI Engineering lane of the AI-SDLC in Figure 2, the prototype is developed in the sub-process Machine Learning and optimized iteratively. The amount of time for the design, implementation and testing of the TRL3 prototype was quite reasonable and the return on investment of an AI-based solution was feasible even with limited resources and prior knowledge about machine learning.

In order to collect all the necessary data, the arcVision Scanner was integrated into the operational environment respectively into the regular daily time schedule for a few days. Fortunately, the recorded data was viable for the validation and testing of the arcVision Model. Eventually, the arcVision System had an adequate accuracy in arc ignition detection and TRL6 could be achieved. However, the model has still some problems with false positives and predicts arc ignitions, which did not exist on the image.

Figure 5 shows various characteristics of arc ignitions, which were part of the data-set and provided valuable insights about the arc ignition patterns, which were out of reach before.

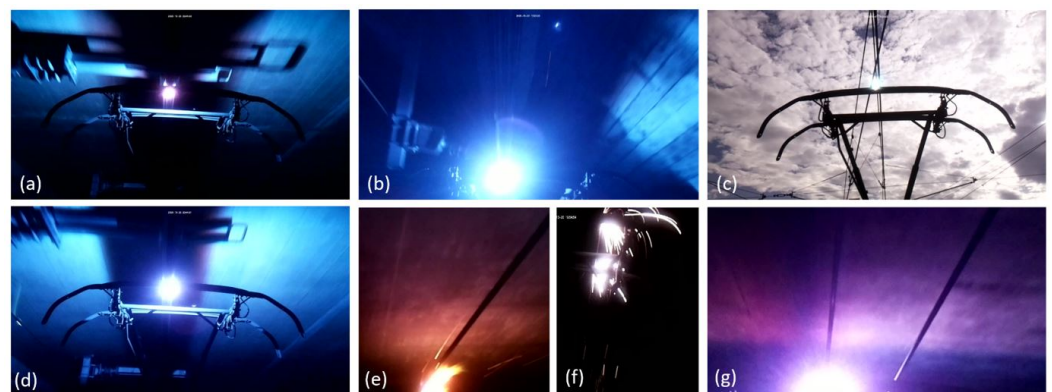


Figure 5. Arc ignitions in tunnels (a,b,d,e,g) and in open air (c,f).

Although, one could argue, that the arcVision Scanner and Model are both demonstrated in the operational environment, the entire arcVision does not yet achieve TRL7, since it is not yet embedded into the overall IT system architecture and neither the data management, nor the model management, nor the rules are defined, for which value ranges of accuracy or F-scores of the arc ignition patterns should trigger a real maintenance task. However, a side effect of the presentation of the prototype and the data-set was the increased awareness of applicability of AI-related technology in the management.

4.3. AI in Rail-Inspect at SBB

SBB is an early adopter when it comes to the application of AI. SBB's goals are to increase the safety of its railways infrastructure, reduce costs, enable new business models and cope with the ever-growing transportation demand. SBB has different use-cases ranging including supervised ML, reinforcement and unsupervised learning, knowledge representation or optimization. Some of them are described on the flatland-challenge website [21]. Below we will provide a summary of one of the flagship cases, called AISI (AI

in Railwaysinspect), which is the first major operational project at SBB that makes massive use of deep learning technologies and can be classified as a TRL9 application.

We also summarize the lessons learned both during implementation and operation. The project AISI was initiated with the aim to automate the track inspection and to close the gap between demand and availability of resources. The goal of AISI is to analyze image data with the objectives to increase the network safety and to reduce the overhead for track inspection. It uses state-of-the-art Deep Learning methods to detect infrastructure defects from high resolution images captured by line-scan cameras of diagnosis vehicles. Five parallel installed line-scan cameras capture detailed images from the tracks and transfer them to the AISI ML system for further analysis. The defects found are marked with bounding boxes and categorized by the type of defects, as depicted in Figure 6 below.

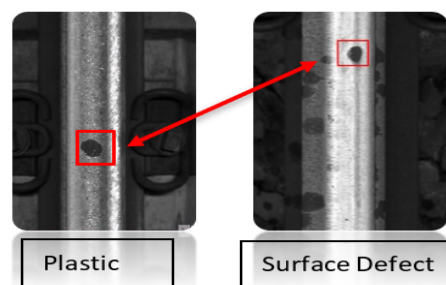


Figure 6. Difficulty in detecting defects based on image data with AISI.

The first ML-model has been put into production but covers only a subset of the 250 different known defect types. However, it achieves much higher detection rates compared to previous non-ML algorithms. AISI was started in 2017 as a research prototype showing promising results. However, the path from a research prototype on TRL4 to an enterprise ready product on TRL9 is very elaborate and expensive, especially for products that have to incorporate many other aspects apart from the yet complex model development. The following sections describe the main challenges to integrate ML components into operational processes and propose possible approaches to tackle them.

4.3.1. Data and Model Life-Cycle

Data is most critical for the success of ML-models and needs to be proactively managed during its entire life-cycle, from its acquisition to its incorporation as part of the model development.

Figure 7 shows a section of the SDLC of Figure 2, which illustrates the image life cycle. It is not only the quantity, but also the quality of data which needs to be managed and the transition from big data to good data is crucial [22]. There are two main issues related to data quality: (1) non-consolidated definition of objects of interest and (2) inconsistent labelling. Usually data is labelled by different experts. Without a systematic business consolidation we observed quite often very heterogeneous definitions of the objects of interest. As depicted in Figure 8, two experts can have a different view about the size of the defects to categorize an object as being a defect or not.

Such different views consequently not only lead to low quality models but also to different opinions from experts, i.e., some will complain that not all defects were detected (false-negatives) and some will complain that too many defects were detected (false-positives). The consolidation is a crucial prerequisite for successful models. The consistency of the labels has also a great impact on the model performance. Quite often we observed many inconsistent labelling, which include too large or too narrow bounding boxes as well as labelling the same defect more than once (Figure 9).

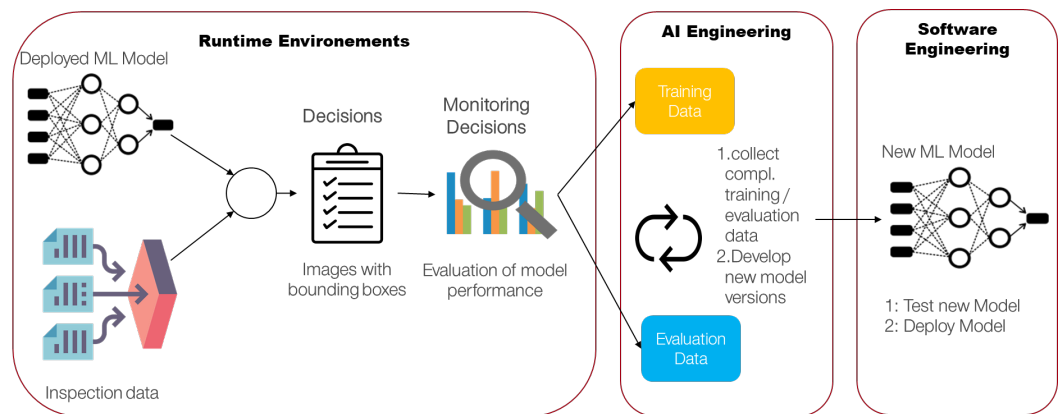


Figure 7. Model and Data Life-Cycle in the SDLC for TRL9 solutions.

Length of defect in mm	Expert 1	Expert 2	
0.7	1	1	✓
0.5	1	0	✗
0.3	1	1	✓
0.25	0	0	✓
0.2	0	1	✗
0.1	0	0	✓

Figure 8. A non-consolidated definition of objects of interest.

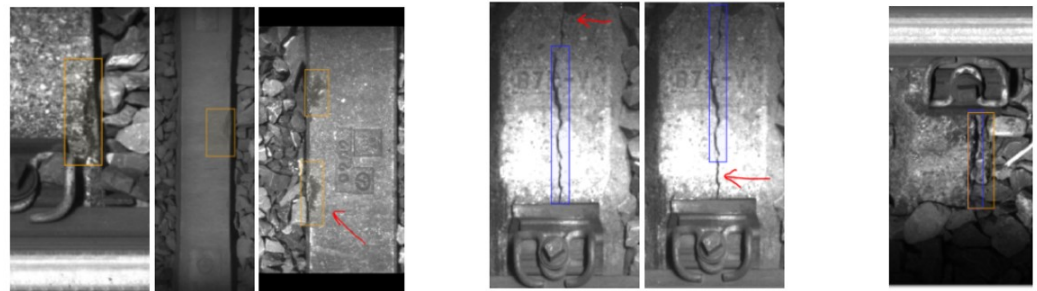


Figure 9. Inconsistent labelling—inconsistent size of boxes confusing the ML model.

The experts need clear guidelines on how to label the data and must be aware of the consequences of inconsistent labelling on model performance. Moreover, the tedious labelling is boring for humans and concentration quickly drops off [10].

Assessing the quality of data and model performance are not one-shot activities, but need to be continuously monitored and managed during their entire life-cycle. As measurement data are inferred with the deployed models, their performance needs to be assessed and based on these results, additional data might be necessary (Figure 7). After that, the models need to be retrained in order to close the gaps identified by the assessments of the model performance. The following reasons lead to the necessity of this process. First, there might be a data and concept shift [23,24] causing different data distributions as well as different mappings of data to labels over time. In our case, if we change the cameras, the resolution of images will most probably increase and consequently trigger a change to consider even smaller defects with a length of 1mm (see Figure 8) to be reported. In the first case, the data distribution has changed, whereas in the second case the mapping of data to labels is different. Furthermore, in many cases, existing data does not cover the entire target application domain or problem space. In AISI, we initially only covered a subset of the entire network. As we rolled-out the model to more tracks we faced new defect situations, which were not covered by the initial model and caused a performance degradation of the

model on these extended network tracks. As depicted in Figure 2, such boundary events caused by drifts in data or changing rules can trigger additional investigations in the ML as well as KE sub-processes to constantly adapt to the changing business requirements. This is the reason why we introduced two complementary procedures, which we call crowd-evaluation and expert-labelling. Both procedures make the data a first-class citizen [25] and support the aforementioned continuous management process.

The goal of crowd-evaluation is to let different experts label data based on the model performance in a crowd-manner (assign the task to more than one expert). As part of the expert-labelling we choose a set of data to be verified on the field, which will be used for the evaluation of the new model version (see Figure 7). Crowd-evaluation aims at increasing the quality of training data at low cost and is part of the consolidation process (Figure 8). Expert-labelling strives towards high quality data, which is used for the performance assessment of the models. Clearly, the amount of data collected by expert-labelling is much lower compared to that of training data. One of the main open issues is the choice of the most representative tracks that should be used for expert-labelling and also the development of technical solutions to ensure that we don't mix tracks that were *team-labeled* with those of *expert-labelling* since training and test data must be disjoint.

4.3.2. Pipelines and Release-Management

A pipeline consists of all steps from data acquisition to data analysis and the necessary post-processing to deliver the data to the experts in the desired format. We have two types of pipelines: (1) a training pipeline, which is used for (re-)training our models and (2) an inferencing pipeline, which is used for analyzing production data from the inspection runs. Both pipelines must be highly automated and traceable. Especially for fulfilling the traceability requirement, it is crucial to version the code, data and the model. To allow this traceability, a well established ML meta data management is essential, that allows for tracking all images that have been used to train a given ML model.

Thus, a well-defined release management needs to be established which incorporates code, data, models and the ML meta data. New models can only be deployed if they are carefully tested beforehand and if it can be ensured, that on one hand the performance of the current models can be assured, and on the other hand, the models infer also correctly on new data. This requires a continuous monitoring and means of automatically evaluating new model versions based on clearly defined evaluation data (see expert labelling).

5. Conclusions and Outlook

There is a considerable amount and broad spectrum of research about the potential of using AI-based IT-systems for various tasks in transportation industry. Many organizations already started to build more "intelligent" systems for autonomous driving, passenger safety, crime detection, traffic management or to implement condition based and predictive maintenance. Further use cases consider building digital twins of train stations based on symbolic AI in terms of analysis of different application cases, case-based reasoning to develop rules from existing paper based customer documents, system reports or to apply sub-symbolic supervised or semi-supervised ML for video-based analysis using the many cameras that are already installed in many public areas. However, many AI use-cases fail due to the high complexity in achieving enterprise-readiness, i.e. in the transition from initial prototypes to validated and proven products. We analysed the different use-cases from three different organizations and provide a BPMN diagram, that visualizes from a behavioural and responsibility perspective, how an AI-based software development cycle can be implemented in the organization. Furthermore, we provide a template of a 2-dimensional domain specific AI landscape, that can provide a good overview of existing AI components with respect to the AI Domains, that have been determined by several technical reports of the European Commission. Based on that, we described different use cases of AI-based components, how they are positioned in the different organizations and where they are located on the proposed AI landscape. We classified the different

AI components according to their maturity levels and described the challenges that are involved during the engineering from initial prototypes to production ready deployable solutions. These cases help to draw some conclusions and provide a shared vision, to establish a common IT Architecture and to work out concepts how the existing AI-based components from different technology readiness levels can be shared and reused among different organizations in order to leverage their potential for all transportation services in Switzerland or even in cooperation with the EU. A common IT architecture will ease the sharing of knowledge, data and models and speed up the adoption of AI in enterprises of a common business domains.

Author Contributions: Conceptualization, S.J.; methodology, S.J., I.F., A.R., D.M. and M.P.; writing—original draft preparation, S.J.; writing—review and editing, S.J., I.F., A.R., D.M. and M.P.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Webb, G.I.; Lee, L.K.; Goethals, B.; Petitjean, F. Analyzing concept drift and shift from sample data. *Data Min. Knowl. Discov.* **2018**, *32*, 1179–1199. [CrossRef]
- Object Management Group. *Business Process Model and Notation*; Object Management Group: Needham, MA, USA, 2021.
- Decision Model and Notation DMN v.1.3, 2021. Available online: <https://www.omg.org/spec/DMN/1.3/About-DMN/> (accessed on 17 December 2021)
- Weber, C.V.; Curtis, B.; Chrissis, M.B. Capability Maturity Model, Version 1.1. *IEEE Softw.* **1993**, *10*, 18–27. [CrossRef]
- Lenarduzzi, V.; Lomio, F.; Moreschini, S.; Taibi, D.; Tamburri, D.A. *Software Quality for AI: Where We Are Now?* Springer International Publishing: Berlin/Heidelberg, Germany, 2021; Volume 404, pp. 43–53. [CrossRef]
- Gartner. The-Cio-Guide-to-Artificial-Intelligence. 2021. Available online: <https://www.gartner.com> (accessed on 17 December 2021).
- Little, C.; McCormick, J. *Measure Your Digital Intelligence Maturity*; Forrester Research: Cambridge, MA, USA, 2019; p. 11.
- Amershi, S.; Begel, A.; Bird, C.; DeLine, R.; Gall, H.; Kamar, E.; Nagappan, N.; Nushi, B.; Zimmermann, T. Software Engineering for Machine Learning: A Case Study. In Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP 2019), Montreal, QC, Canada, 25–31 May 2019; pp. 291–300. [CrossRef]
- Samoili, S.; López Cobo, M.; Gómez, E.; De Prato, G.; Martínez-Plumed, F.; Delipetrev, B. *AI Watch—Defining Artificial Intelligence. Towards An Operational Definition and Taxonomy of Artificial Intelligence*; EPrints: Southampton, UK, 2020; pp. 1–90. [CrossRef]
- Martínez-Plumed, F.; Gómez, E.; Hernández-Orallo, J. *AI Watch: Assessing Technology Readiness Levels for Artificial Intelligence*; Technical Report; Joint Research Centre: Ibaraki, Japan, 2020. [CrossRef]
- Jöhnk, J.; Weißert, M.; Wyrski, K. Ready or Not, AI Comes—An Interview Study of Organizational AI Readiness Factors. *Bus. Inf. Syst. Eng.* **2021**, *63*, 5–20. [CrossRef]
- Francesco, C. *AI knowledge Map: How to Classify AI Technologies, a Sketch of a New AI Technology Landscape*; Springer: Berlin/Heidelberg, Germany, 2018.
- Morley, J.; Machado, C.C.; Burr, C.; Cowls, J.; Joshi, I.; Taddeo, M.; Floridi, L. The Ethics of AI in Health Care: A Mapping Review. *Philos. Stud. Ser.* **2021**, *144*, 313–346. [CrossRef]
- Floridi, L. What the Near Future of Artificial Intelligence Could Be. *Philos. Technol.* **2019**, *32*, 1–15. [CrossRef]
- Monitoring, I. *Rail Transport Research and Innovation in Europe*; Publications Office of the European Union: Luxembourg, 2021. [CrossRef]
- Schroeder, C. *The Challenges of Industry 4.0 for Small and Medium-Sized Enterprises*; Friedrich Ebert Foundation: Bonn, Germany, 2015; pp. 1–28.
- Hinkelmann, K.; Martin, A.; Fill, H.-G.; Gerber, A.; Lenat, D.; Stolle, R.F.v.H.E. Combining Machine Learning and Knowledge Engineering. In Proceedings of the AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering (AAAI-MAKE 2021), Palo Alto, CA, USA, 22–24 March 2021.
- Peraic, M. Feasibility Study: AI-based Passenger Counting System for Public Transit. Bachelor’s Thesis, University of Applied Sciences and Arts Northwestern Switzerland, Basel, Switzerland, 2019.

19. Jüngling, S.; Peraic, M.; Martin, A. Towards AI-based Solutions in the System Development Lifecycle. In Proceedings of the AAAI 2020 Spring Symposium on Combining Machine Learning and Knowledge Engineering in Practice (AAAI-MAKE 2020), Palo Alto, CA, USA, 23–25 March 2020; Volume I, p. 2600.
20. Morandi, D.; Jüngling, S. Anomaly Detection in Railway Infrastructure. In Proceedings of the AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering in Practice (AAAI-MAKE 2021), Palo Alto, CA, USA, 22–24 March 2021.
21. SBB. Flatland-Challenge. Available online: <https://www.aicrowd.com> (accessed on 17 December 2021)
22. Miranda, L.J. Towards Data-Centric Machine Learning: A Short Review. Available online: <https://lvmiranda921.github.io> (accessed on 17 December 2021).
23. Tian, J.; Hsu, Y.C.; Shen, Y.; Jin, H.; Kira, Z. Exploring Covariate and Concept Shift for Detection and Calibration of Out-of-Distribution Data. *arXiv* **2021**, arXiv:2110.15231.
24. Wang, X.; Wang, Z.; Shao, W.; Jia, C.; Li, X. Explaining Concept Drift of Deep Learning Models. In *Cyberspace Safety and Security*; Series Title: Lecture Notes in Computer Science; Vaidya, J., Zhang, X., Li, J., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; Volume 11983, pp. 524–534. [[CrossRef](#)]
25. Ng, A. A Chat with Andrew on MLOps: From Model-centric to Data-centric AI. Available online: <https://www.deeplearning.ai/wp-content/uploads/2021/06/MLOps-From-Model-centric-to-Data-centric-AI.pdf> (accessed on 17 December 2021).