DEPARTMENT OF INFORMATICS

TECHNISCHE UNIVERSITÄT MÜNCHEN

Seminar Paper

# The EU's AI Act Proposal: Technical Documentation and Discussion of KONUX Switch

**Ozan Aydin, Andreas Heckl, Janick Hofer, Suchithra, Sophie Witt**

We confirm that this seminar paper is our own work and we have documented all sources and material used.

Munich, 18.06.2023 Ozan Aydin, Andreas Heckl, Janick Hofer, Suchithra, Sophie Witt

# Contents

# 1 Regulation of KONUX Switch [Suchithra]

## 1.1 Introduction

KONUX was founded in Munich in 2014 by Andreas Kunze, Vlad Lata, Max Hasler, and Dennis Humhal to enable the digitization of the rail industry. They work with Deutsche Bahn, one of their main business partners, on implementing Infrastructure 4.0. In 2017, KONUX ramped up its Series A to $16M, in order to expand their Artifical Intelligence (AI) capabilities and expand on the European market [1]. In the fall of 2019, they started working with three new European customers and are now active in seven countries worldwide. Their customer base was raised to 10 international customers in Europe and Asia. In the winter of 2020, KONUX won the Deutsche Bahn tender for the digitization of switches [1].

KONUX is the first AI scale-up in railway. It has been known for next-generation predictive maintenance, network usage, and traffic monitoring and planning solutions for railway infrastructure management. It provides three solutions:

- KONUX Switch: a rail switch predictive maintenance system that employs machine learning and the Industrial Internet of Things to offer delay-free switches at an optimal cost;
- KONUX Network: an application for use tracking and inspection planning that offers a more precise view of network traffic and its effects on the infrastructure; and
- KONUX Traffic: a method for schedule optimization and delay reduction that aids route directors, dispatchers, and planners in overcoming capacity constraints and improving scheduling

.

KONUX Switch is the one that we have considered to evaluate the applicability of the European Union Artificial Intelligence Act (the "EU AI Act"). In our pursuit of examining responsible technology and ensuring compliance with the EU AI Act, KONUX Switch came across as an exemplary embodiment of innovation and consideration for the betterment of society. This AI scale-up holds a primary focus on revolutionizing rail transport toward a sustainable future. Moreover, it distinguishes itself as a trailblazer by offering tangible solutions for predictive maintenance, network optimization, traffic monitoring, and comprehensive railway infrastructure management planning.

With its innovative implementations, KONUX Switch aims to revolutionize the very essence of rail transportation, paving the way for a sustainable future. The impact of their efforts extends far beyond the domains of technology, encompassing society as a whole and public transportation in particular. By reimagining the management and maintenance of railways, KONUX Switch is reshaping the landscape of public transportation, ensuring efficiency, safety, and environmental consciousness.

By using KONUX Switch as the focal point of our discussion, we hope to cast light on the potential for responsible technology to transform not only industries but also the lives of those who rely on public transportation. Their solutions serve as a beacon, directing us toward a future in which technology and responsible practices coexist, harmonizing progress with ethical considerations. So an evaluation of such a technology in consideration of the emerging Act is essential both at the technology level and societal usage level.

## 1.2 Scope and Applicability

### 1.2.1 Territorial Jurisdiction

Regarding the AI system's geographical area of application, regardless of where the supplier or user is located, the main concern is whether the system's effects are felt in the European Union (EU). This might result in the EU AI Act being applied extraterritorially in a way that goes much beyond the boundaries of the EU. According to Article 2 of the EU AI Act, it will be applicable to:

(a) Providers that first supply commercially (whether or not free of charge) or put an AI system into service in the EU (putting into service involves making it available for use by a "user" or for use by the provider itself), regardless of whether the providers are located inside or outside the EU

(b) Users of AI located in the EU; and

(c) Providers and users located outside the EU, if the output produced by the system is used in the EU.

Since KONUX was founded in Germany and is putting an AI system into service in the EU, the EU AI Act is applicable to KONUX based on territorial jurisdiction. With regard to its territorial scope of application, the focal point is whether the impact of the AI system occurs within the EU, regardless of the location of the provider or user. Since KONUX Switch is used within the territorial scope of the EU, this would be bound by EU law.

### 1.2.2 Subject-Matter Jurisdiction: Risk Level

**1) Can KONUX Switch be classified as high-risk?**
Two aspects should be in consideration here before listing if KONUX Switch falls into the category of High-Risk AI. First, we have done the "examination of law" pertaining to High-Risk AI and secondly, we have conducted the examination of legal provisions pertaining to "critical infrastructure" as per the EU AI Act.

We have examined the applicability of the law as follows: The EU AI Act defines High-Risk AI as herein below (Article 6):

- Irrespective of whether an AI system is placed on the market or put into service independently from the products referred to in points (a) and (b), that AI system shall be considered High-Risk where both of the following conditions are fulfilled:
  (a) The AI system is intended to be used as a safety component of a product, or is itself a product, covered by the Union harmonization legislation listed in Annex II;
  (b) The product whose safety component is the AI system, or the AI system itself as a product, is required to undergo a third-party conformity assessment with a view to the placing on the market or putting into service of that product pursuant to the Union harmonization legislation listed in Annex II.
- In addition to the High-Risk AI systems referred to in paragraphs a and b above, AI systems referred to in Annex III shall also be considered High-Risk.

We have examined the legal provisions of critical infrastructure as follows:
The High-Risk AI has been categorized in Annex III further as follows:

- **Critical infrastructures (e.g. transport)**, that could put the life and health of citizens at risk;
- Educational or vocational training;
- Safety components of products (e.g. AI application in robot-assisted surgery);
- Employment, workers management and access to self-employment;
- Essential private and public services ;
- Law enforcement;
- Migration, asylum and border control management ;
- Administration of justice and democratic processes;
- Surveillance systems.

KONUX has created sensors that businesses may employ to increase the productivity of their own equipment. For this aim, KONUX designed a discrete gray box that is fitted to the switches of Deutsche Bahn, for instance. It has an advanced sensor system that detects vibrations and helps the railroads better manage their own train network while preventing breakdowns and delays. It provides self-learning algorithms in cloud-based applications that provide accurate data evaluation.
The EU AI Act clearly explains its objective and rationale, stating that AI systems intended for use as safety components in the management and operation of road traffic, as well as the supply of water, gas, heating, and electricity, are appropriately classified as High-Risk. This

classification is justified by the potential risks posed to the life and health of individuals on a large scale, as well as the significant disruptions that can occur in the normal functioning of social and economic activities in the event of failure or malfunctioning.
Based on the classification given under the EU AI Act and the AI that KONUX uses, it is safe to say that KONUX Switch falls within the ambit of High-Risk AI.

**2) Is KONUX Switch categorized under critical infrastructure?**
Since the EU AI Act categorizes the applications related to transport, education, employment and welfare, among others, as high-risk AI, the tool falls within that ambit as it is deployed in transportation, one of the critical infrastructures.

**3) Should KONUX Switch be banned?**
Title III, Article 5 lists the AI practices that should be banned. Depending on the degree of danger that the AI may produce, the guidelines (Article 5) define duties for providers and consumers. Systems that use subliminal or intentionally manipulative tactics, take advantage of people's vulnerabilities or are used for social scoring (classifying people based on their social behavior, socioeconomic status, or personal characteristics) would all be strictly prohibited. The EU AI Act now prohibits the use of AI systems in the following ways:

i. "Post" remote biometric identification systems, with the exception of law enforcement for the prosecution of significant offenses and only after court authorization; "Real-time" remote biometric identification systems in publicly accessible locations.

ii. Predictive policing systems (based on profiling, location, or prior criminal behavior); Biometric categorization systems utilizing sensitive factors (e.g. gender, race, ethnicity, citizenship status, religion, political orientation);

iii. Emotion recognition systems in workplaces, educational institutions, border management, and law enforcement; indiscriminate biometric data scraping from social media or CCTV video to generate face recognition databases (violating privacy and human rights).

Evaluating the above provisions and the lists that are given under Article 5, it is evident that KONUX Switch does not fall into any of these AI practices. KONUX Switch does not use a biometric identification system, or it is not used in the field of predictive policing systems or does not involve an emotion recognition system; hence it should not be banned. KONUX Switch is used only for the purpose that is already allowed by the EU AI Act.

**Conclusion** In conclusion, we can state that the specific KONUX Switch application that we have examined for this study, a prominent AI utilized in railway transportation and operating within Germany, unmistakably falls within the scope of the EU AI Act's category of High-Risk AI, explicitly pertaining to critical infrastructure. As an essential component of the railway sector, KONUX Switch's potential impact on the safety, reliability, and efficiency of critical infrastructure necessitates its inclusion within this regulatory framework. By recognizing

KONUX Switch's high-risk nature, the EU AI Act ensures the implementation of stringent measures to mitigate potential hazards, safeguard public interests, and secure the responsible development and deployment of this technology in railway transportation systems. At the same time, KONUX Switch does not fall within the prohibited list of the EU AI Act.

# 2 Technical Documentation According to Art. 18, Art. 11 and Annex IV

The following is a technical documentation for the KONUX Switch application, as described in Chapter 1.

According to Article 18, providers of high-risk AI systems, which KONUX Switch is categorized as according to the reasoning in Chapter 1, are required to provide technical documentation. Article 11 specifies that this documentation shall be drawn up before that system is placed on the market or put into service and shall be kept up-to date.

However, the exact scope of the technical documentation is described in Annex IV. The following chapter will guide chronologically through the structure of the Annex. However, to ensure a sufficient scope of the documentation, only Art. 13 and Art. 15 of the EU AI Act will be dealt with in detail in the following documentation. All other articles required for the technical documentation will be dealt with partially, if necessary.

## 2.1 General Description of the System [Ozan Aydin]

### 2.1.1 Overview

KONUX Switch is an industrial, predictive maintenance system for railway switches. The intended purpose of KONUX Switch is to use Internet of Things (IoT) devices and AI to improve network capacity, extend asset lifetime and reduce costs.

The application relies on data that are coming from the sensors which are installed within the tracks. The sensors are tested under industrial workload and pressure to ensure they can withstand various environmental factors in the field. These sensors feed the KONUX Switch backend continuously with various data, taking into consideration factors like weather, track pressure, soil movement and so forth.

With the aforementioned data in hand, KONUX Switch utilizes machine learning models to apply predictive maintenance, basically tracking the sensor data and predicting if and when switches need to be maintained, allowing organizations to prepare and utilize their resources beforehand.

The following parts of the documentation contains a number of technical jargon, which will be briefly described here:

- **Switch**: Switches consist of many different parts and are used to allow trains to move from one track to another.

- **Frog**: Part of the switch, located at the very end of the structure. Frogs enable train wheels to transport from one track to another.
- **Trackbed**: Refers to the ground on which the railroad is laid.

### 2.1.2 Version

Latest Stable Version: 1.23.0
Date: 23.02.2023

### 2.1.3 The Team Behind KONUX Switch

Beyond the founders, the current KONUX Switch team is lead by Thomas Böhm and David Kronmüller. Also part of the team is Arthur van de Wiele as the Head of Embedded Systems department, Olga Spackova as the Analytics Team Lead, and Martin Mason as Director of Productive Maintenance.

### 2.1.4 Interactions with Different Systems

As mentioned in the Overview, KONUX Switch relies on data that is coming from special IoT devices installed by KONUX. The data is then processed by the backend algorithm and the corresponding results are displayed into the dashboard. In this state, neither the data nor the algorithm is open-source, which heavily restricts any outside interactions with different systems.

The most crucial interaction is between the data and Amazon AWS systems. KONUX Switch uses remove EC2 servers that are provided by AWS to run machine learning models, as running these services locally would take up immense amounts of resources. That means that the data gathered by the sensors leave the local pipeline and interact with AWS's system.

Overall, the intended purpose of KONUX Switch is not to appeal to the general public but rather very specific to the welfare of a certain hardware in an industrial setting, which leads to the conclusion that apart from the intended end users, remote servers that are provided by AWS and hardware components, there will be no interactions with different systems.

### 2.1.5 Market Release

The possible end users for KONUX Switch are either the governmental bodies or the private companies that own and maintain the railroads, therefore the AI system is not available for the general public to consume. When available, KONUX Switch includes both hardware and software components. Necessary manpower is provided by the company to insert the required hardware into the tracks, and after the setup is complete, the end-users gain access to the dashboard where they can see the human-readable results of the machine learning models. The models themselves, however, are in a black box format and the API for the models are not included in the market release.

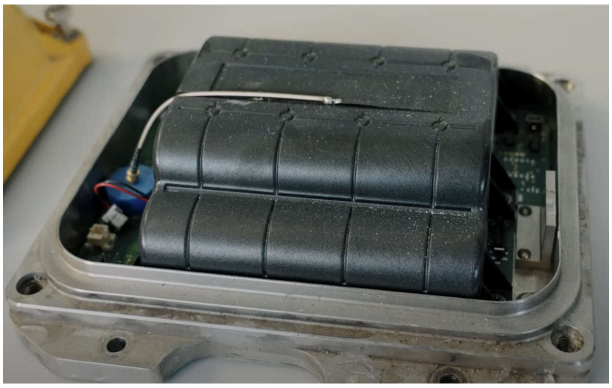

Figure 2.1: The figure shows the custom made KONUX Switch sensors, with its internal protective casing attached. [2]

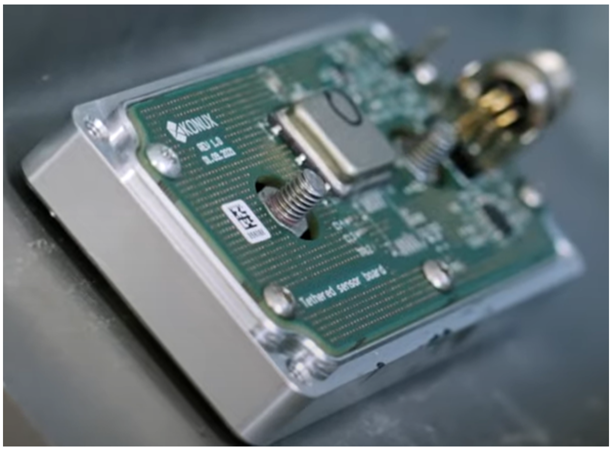

Figure 2.2: The figure shows the circuit of the sensors. The casing is visible under the protective casing, which has been removed in the picture. [2]

### 2.1.6 Hardware Specifications

The final market release of KONUX Switch is intended to run on any computer, smartphone or tablet. The actual machine learning models themselves are running on Amazon servers, more specifically AWS EC2 instances.

### 2.1.7 Components

As mentioned before, the AI system consists of two components, hardware and software. The hardware component consists of custom made sensors, with hard plastic shells that cover the motherboard within, which can be seen in Figure 2.1. Inside the shell, the body of the sensors are located 2.2, and an actual application of the sensor in the field, within the railroad tracks can be seen in Figure 2.3. All together, the custom made sensors are very durable and have a battery life of four years.

Figure 2.3: The figure shows the final form of the KONUX Switch sensor, installed within the trackbeds. The external yellow protective casing is also visible, which further protects the sensors from any damage from outside. [2]

### 2.1.8 User Instructions

End users will be able to interact with the AI system through the KONUX Switch dashboard. The dashboard is available through the KONUX app, which the user has to log in with their credentials. Once logged in, the application opens with the main dashboard, which can be seen in Figure 2.4.

Through the main page, all necessary information can be accessed. End users can see the overall health of their switches or can search for a switch using the search bar on the top right corner. Users can also see two of the most critical infrastructure errors on this page, mainly vertical displacements and problematic frogs.

Next to the "Dashboard" tab, users can click on "My Region" to view all the trackbed, frogs and load factors that belong to a specific region, as visible in Figure 2.8. Users can choose to view detailed information about the frog and trackbed conditions or load factors related to certain switches in the specific region. A detailed page about the frog conditions show the forecast on the frogs health, past maintenance work done on the frog and a visualization which shows the past and the possible future health of the frog. A detailed page about the trackbed condition visualizes the vertical displacement of the trackbed and the history of maintenance work, while also showing predictions of further displacement in the future, both in written form and in graph form. Users can also view load factors of each switch, which show a special KONUX Switch metric that unifies tonnage, number of trains and speed. Users can view multiple facts about the traffic, including number of trains that pass from that switch, the average weight, the average speed and so on.

Next to the "My Region" tab, "Maintenance" tab displays all the information about the maintenance history, as shown in Figure 2.6. Users can filter the information displayed in the screen by region, time, type of maintenance and evaluation parameters. On the top left corner, users can see the visualization of the effectiveness of the previous maintenance work.

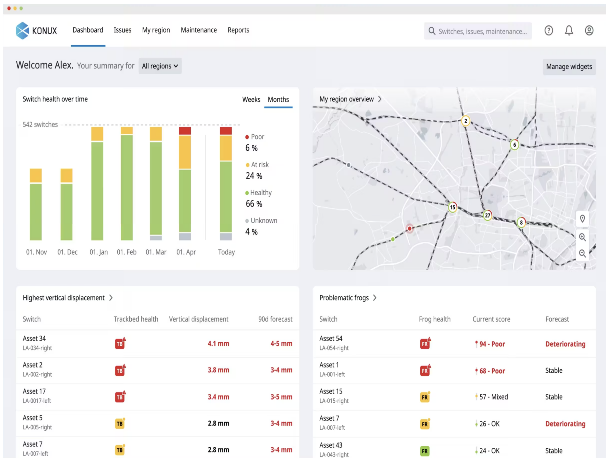


Figure 2.4: The figure shows the main dashboard when you open the KONUX Switch app. On the top, there are tabs that a user can navigate to, and in the main screen, users are able to see the overall health of the switches as well as the map overview. Also visible are the trackbed and frog conditions, visible in the bottom of the page. [2]

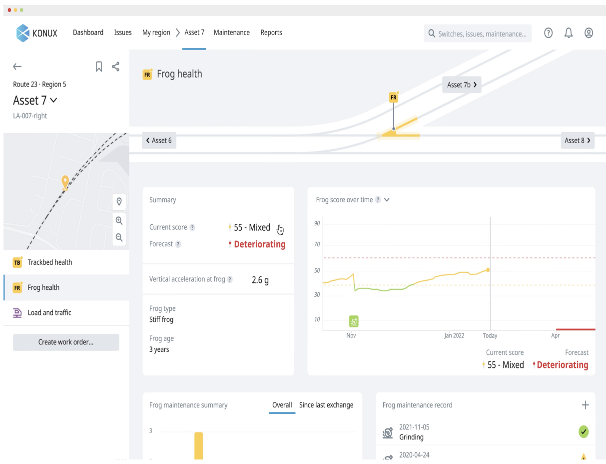


Figure 2.5: The figure shows the frog health information screen for a single switch. Users can see where the overview on the map, the prediction for the condition of the frog and past maintenance work that was done on the frog. [2]

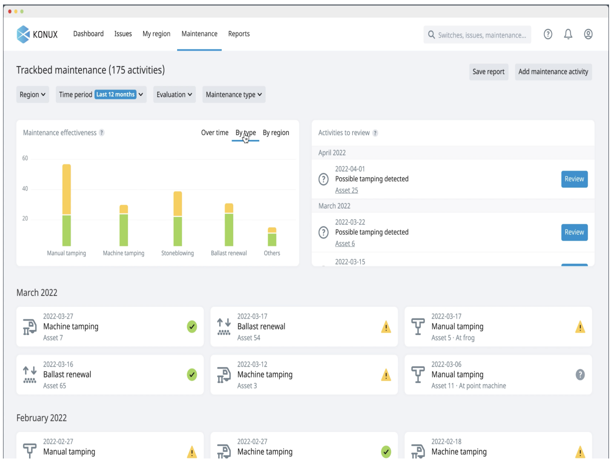


Figure 2.6: The figure shows the past maintenance page. In more detail, it is possible to see the past maintenance activities and their effects on the condition of the trackbed in the graph, while on the bottom of the screen individual maintenance work that has been previously done on the trackbed can be seen. [2]

Scrolling down, users are able to view every single maintenance work done over the previous months, with information displaying which time and the name of the switch. Clicking on these panels will display a more detailed information about the maintenance work. On the top right side of the screen, users can view recent activities that KONUX Switch might have caught, like possible tamping operations near certain switches.

## 2.2 Design Specifications and Main Classification Choices [Sophie Witt]

This paragraph explains the design specifications of KONUX Switch, the assumptions that were made and how the target values, i.e. the classes for the classification task, were defined.

KONUX Switch uses an approach that calculates the Remaining useful life (RUL) which is essentially the time that is left before a switch fails. However, train companies still need time to perform switch changes to prevent them to fail before the maintenance is performed. Thus, switch failures need to be revealed between 3 and 5 days in advance. Subsequently, the prediction of the RUL can be understood as a classification in two classes: one contains all examples with a RUL smaller than 5 days and the other the remaining ones, implying a RUL of 5 days or more [3]. The two explained classes essentially represent the displayed forecast labels "stable" or "deteriorating" which are shown in the KONUX Switch application (see figure 2.8 in chapter 2.3.1).

Of course, the application displays much more data about the status of the switches to the user, including for instance the vertical displacement (see 2.7) and the current score (see 2.8). However, the ultimate goal is to predict the need for maintenance using the RUL and the explained classification.

In order to be able to predict the RUL, and thus classify the switches as mentioned, several parameters have to be taken into account. The electric power consumption of the engine is an indicator for the RUL of a switch. Furthermore, the RUL can dependent on the temperature, humidity and the switch construction characteristics, e.g. single switches are less complex than curve or double switches. KONUX performs supervised learning to train the model. Thus, recorded switch failures are needed as a condition reference. Their training data contains all mentioned features and the recorded failures including the failed parts and the reason. Additionally, it comprises information about the performed maintenance is needed as this clearly influences the prediction of the RUL [3].

After a heavy research and development phase, the KONUX team decided to select a Support Vector Machine (SVM) [1] to perform the predictive maintenance task. Generally, a SVM tries to find a hyperplane in the feature space divides the classes the best by maximizing the distance between data points. As the RUL classes are not linear separable, they make use of a Kernel-Function that transforms the feature space into a higher dimension. Consequently, the development team trains the SVM to find a hyperplane for the resulting feature space. Using this approach, the most important hyperparameter for the SVM is said Kernel function. Following other researchers in the field, KONUX uses a Radial-Base-Function Kernel because it resulted in the best performance results in their experiments [3]. However, their team is continuously improving the model by tuning parameters and constantly evaluating other machine learning methods for the use case in order to be able to provide the best product possible to their customers.

## 2.3 Transparency and Provision of Information [Janick Hofer]

As indicated in the introduction to this chapter, we focus on two articles of the EU AI Act. One of these is Art. 13, which first deals with the transparency of the product and then specifies the information that should be provided to the user.

### 2.3.1 Transparency

As a first step of covering this article, the product is evaluated in terms of transparency according to Art. 13.1. The aim is to prove that the AI system's operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately. These two requirements - interpretability and usability - will be looked at below.

With regard to the interpretability of the output, it is worth analyzing the interfaces of the product. A closer look reveals that predictions are not necessarily marked as AI-based, but that it is clear that these predictions must be based on an algorithm. This is exemplified in

[1]This information about the machine learning model is taken from the cited paper that was published by Thomas Böhm (the current Chief Data Scientist of KONUX) from 2017 due to lack of later publications. Thus, this information could be outdated.

Figure 2.7, where the vertical dividing line signals at which point in time we are currently at. All metrics concerning the future are supported by the label "estimated". The variance of the prediction is indicated by a broadening estimation interval, so it is clear to the user that this is a prediction. In addition, all metrics have a question mark symbol, which gives the user more detailed information about the respective metric when hovering over the symbol. Thus, when analyzing the AI output, it can be summarized that despite the non-direct labeling, there is an appropriate interpretability.


Figure 2.7: The figure shows a screenshot of the KONUX Switch product. In more detail, one sees a graph showing the vertical displacement in the past and predicted for the future. In addition, percentage value for exceeding the critical 3mm is given for future values. [2]

In addition to the interpretability of the output, the usability of the data must be evaluated. For this purpose, KONUX Switch provides overview tables that bundle the different predictions and metrics in one summary. Based on these different metrics, the user is then given a forecast in the form of a clear categorization into "Stable" or "Deteriorating", so that he or she does not have to draw an own conclusions from the data (see Figure 2.8). This processing of the data bundled with individual graphs for better visualization of the individual metrics ensures sufficient usability of the output. KONUX's general awareness of interpretability and usability is also shown in their Spring 2022 release announcement, in which they state "The improved data visualization is optimized for AI insights to reflect how predictive AI models work and enable users to intuitively interpret and utilize data intuitively. For example, the prediction features now provide a range of forecast values indicating differing probabilities, instead of visualizing specific estimates subject to frequent daily changes. This way, asset owners can plan their actions flexibly based on varying chances of an issue." [4]

In addition to the argumentation just made, Art. 13.1. refers to Chapter IV, more precisely

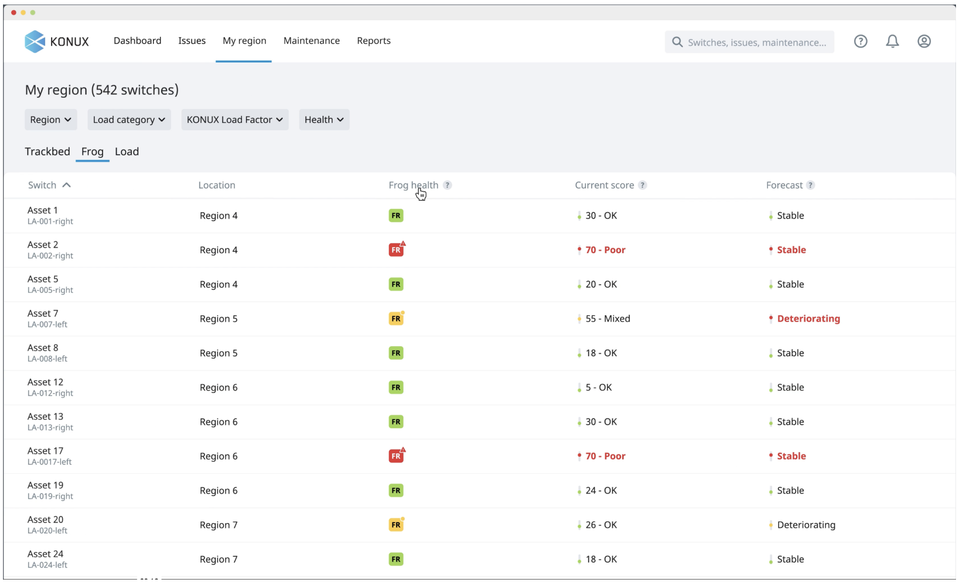


Figure 2.8: The figure shows a screenshot of the KONUX Switch user interface. It displays the summary table that combines the different metrics for the different monitored switches into one overview. From the metrics, a forecast estimate is offered by KONUX, which includes recommendations for action. Additionally, the table can be filtered and sorted as needed. [2]

to Art. 52, which deals with "Transparency obligations for certain AI systems". Taking a closer look at this article, Art. 52.1 states that the AI system shall be designed in such a way that a natural person is informed that he or she is interacting with an AI system. A natural person is defined as a human being that is a real living person. The criterion is weakened by stating that this is not necessary if it is obvious by the circumstances and the context of use. Based on the reasoning about interpretability and usability, the circumstances and context of use seem to be obvious since there is a clear separation of data points and predictions due to the clear marking of future values. Additionally, it can generally be stated that KONUX advertises the product with an AI-first approach, so it should be clear to the user that AI-based algorithms are used in the calculation of forecasts. In this respect, KONUX also adheres to the rules with regard to this article. Art 52.2 and Art 52.3 are not relevant for the KONUX product, since the system does not fall under the category of emotion recognition system, biometric categorisation system, or image/audio/video generation and manipulation for resemblance.

In summary, KONUX Switch complies with the requirements for transparency. To remove any last doubts regarding the user's knowledge about the use of AI, we suggest a disclaimer at the bottom of the page indicating that the algorithms are calculated based on AI. For a detailed discussion about the transparency definition, we refer to Chapter 3.1.1.

### 2.3.2 Provision of Information

The following chapter deals with the provision of information, described in detail within Art. 13.2 and Art. 13.3. The first paragraph specifies that the information should be made available online or otherwise include concise, complete, correct and clear information that is relevant, accessible and comprehensible to users. This rather general statement is particularized in Art. 13.3 by listing the required information in detail. In order to get a comprehensive overview over this information, the points specified in the paragraph will be elaborated chronologically in the following. This will lead to cross-references within this documentation, as some points have already been described in detail in other chapters.

As a first step, we lay out the identity and contact details of the provider:
KONUX GmbH, Flößergasse 2, 81369 München, Germany
Email: info@konux.com
Phone: +49 89 18955010

Regarding the characteristics, capabilities and limitations of performance of the AI systems following information can be provided:

- For the intended purpose of the AI system, the reader is referred to Chapter 2.1.1.
- Also the level of accuracy, robustness and cybersecurity is already extensively covered. The reader is encouraged to go through Chapter 2.4.
- With regard to circumstances that may lead to risk to the health and safety or fundamental rights, sufficient safety of the product must be ensured. Otherwise in the case of incorrect analysis of the points, this might lead to serious accidents. Looking at the safety precautions taken by KONUX, the hardware is fully certified since 2019 and meets state-of-the-art security requirements [5]. Furthermore, the AI system has been tested in a well-documented development process, so that the above mentioned safety measures described in Chapter 2.4 are guaranteed. The user of the analysis tool should be defined within the customer company, but since beyond that the rail network is not directly manipulated, but Konux Switch rather intended for remote analysis, no additional risks to health, safety or fundamental rights occur.
- Regarding the performance, insightful information is provided by KONUX's biggest customer, Deutsche Bahn: Of the contract, concluded with them in 2020, for 650 turnouts, the latter certified the KONUX Switch solution with the passed verification of almost 90 percent prediction accuracy for the first part of the requirements [6]. This public data appear to be sufficient evidence of Provision of Information. The disclosure of internal test data has been waived so far, but these can potentially be partially published again if absolutely necessary.
- A precise specification of the input data or about the split of the dataset does not seem necessary, as this was simply obtained from the rail network, but does not deviate from normal practice for collecting an AI dataset. Accordingly, as specified in Art. 13.3(b)(v), there is no need to elaborate on this point.

Regarding changes that will be completed in the near future, a new version of KONUX Switch is planned for fall 2023. The full scope of the new features is not yet decided. As soon as the scope of the update is determined, this information will be provided to the user.

The human oversight measures, described in Art. 14, pursue the goal of reducing risk through human oversight. For this purpose, the system must be supervised by a human during use. A comprehensive documentation for human oversight is out of scope of this report (as only Art. 13 and Art.15 are covered in detail).

The expected lifetime of the hardware is 12 years, to be distinguished from the AI system, whose digital SaaS platform including algorithms is usable for lifetime. However, this assumes that the interface between data collection by the hardware and the software is constantly maintained. In addition, we propose a biennial check of the IoT sensors, which can extend the lifetime of the hardware to 20 years.

The information listed here is accessible to the user online via the various channels, so that the requirement regarding the provision of information is also fulfilled. In our opinion, a bundled form of information can be dispensed with, as all information is easily accessible. However, if stronger measures were required, the scope of such bundling needs to be discussed again.

## 2.4 Accuracy, Robustness and Cybersecurity [Sophie Witt]

This section demonstrates the compliance of KONUX Switch with Article 15 of the EU AI Act. Thus, it illustrates that the system is developed in such way that it achieves in the light of the intended purpose, an appropriate level of accuracy, robustness and cybersecurity.

### 2.4.1 Accuracy

This subsection reasons why KONUX Switch achieves an appropriate level of accuracy for the specific use of the application. Furthermore, it elaborates why the chosen accuracy metric is appropriate for the task performed by the AI system.

The subsequent paragraph explains the setup to test the accuracy of the AI system. As stated above, the intended purpose for KONUX Switch is to predict the need for maintenance, i.e. classify a switch as "maintenance needed" or "no maintenance needed". This binary classification problem can however also be expressed using a multi-class classification to improve customer satisfaction and make the model more precise. For development and testing purposes, the interval boundaries were manually chosen at a RUL of $\{0.125, 1, 3, 5, 14, 28, 61, 122, 244, \infty\}$ days which results in ten distinct classes. In their experiments, they used 184278 data tuples in total. These were collected from March 2007 to March 2009 by Deutsche Bahn. The train/test split follows a widely used standard of a 70/30 train/test split. [3]

One approach to evaluate the performance of KONUX Switch is to visualize it in a confusion matrix. For a binary classification task, the confusion matrix has four quadrants which contain the True Positive, True Negative, False Positive and False Negative rate. In their case, the cells of the 10x10 matrix show the fraction of examples belonging to the real class in the row and being classified as the class indicated by the column. These performance indicators can be used as a basis for the calculation of more specific performance metrics. For the KONUX Switch application, the confusing matrix shown in 2.1 was assessed[2]. It can be noted that misclassifications mainly occured in the area of RUL values below 1 day which does not have a significant impact on the classification into "RUL below 5 days" and "RUL of 5 days or less" that builds the core of the AI system. Additionally, the model achieves have a very high True Positive rate for all classes but the smallest one (RUL of 0.125 days). Thus, the confusion matrix demonstrates that the accuracy of KONUX Switch achieves an appropriate level in the light of the intended binary classification for predictive maintenance.

| | **Hypothetical class [d]** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Real class** | 0.125 | 1 | 3 | 5 | 14 | 28 | 61 | 122 | 244 | 1 |
| **0.125** | 0.824 | 0.088 | 0.000 | 0.000 | 0.004 | 0.021 | 0.034 | 0.017 | 0.000 | 0.013 |
| **1** | 0.024 | 0.952 | 0.017 | 0.000 | 0.000 | 0.001 | 0.004 | 0.003 | 0.000 | 0.000 |
| **3** | 0.000 | 0.012 | 0.976 | 0.008 | 0.000 | 0.001 | 0.001 | 0.000 | 0.001 | 0.001 |
| **5** | 0.000 | 0.000 | 0.015 | 0.974 | 0.010 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 |
| **14** | 0.000 | 0.000 | 0.000 | 0.002 | 0.993 | 0.002 | 0.001 | 0.000 | 0.000 | 0.000 |
| **28** | 0.001 | 0.000 | 0.000 | 0.000 | 0.002 | 0.994 | 0.003 | 0.000 | 0.000 | 0.000 |
| **61** | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.001 | 0.995 | 0.002 | 0.000 | 0.000 |
| **122** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 | 0.995 | 0.002 | 0.000 |
| **244** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 | 0.995 | 0.002 |
| **1** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.999 |

Table 2.1: This is the confusion matrix of the SVM model. Each cell counts the number of examples belonging to the real class in the row and being classified as the class indicated by the column divided by the total number of examples in the real class.

Another approach to evaluate the performance of KONUX Switch is to select an expressive performance metric. For the particular task they are trying to solve, the chosen metric needs to be invariant to the class distribution at hand as it is likely to have many more examples with the ground truth label "no maintenance needed" than with "maintenance needed". The Matthews Correlation Coefficient (MCC) introduced by Matthews in 1975 handles the stated imbalances into account while being good at detecting small changes in the confusion matrix. Thus, it is less sensitive to the data distribution than other metrics [7]. The range of MCC values can vary from -1.0, indicating a completely wrong classification, to 1.0, representing a perfect classification. For the predictive maintenance task performed by KONUX Switch, the current model achieves a $MCC = 0.989$[3] which indicates a very high proportion of correctly

[2]Here we refer to the experiments conducted in the paper by Thomas Böhm (Chief Data Scientist at KONUX) published in 2017 due to the lack of newer public information. Thus, the real performance of KONUX Switch might have improved or another approach is used today. In addition, we focus on the accuracy of the core functionality as explained above.

[3]The presented numbers are again from 2017 and thus might be outdated.

classified examples. In summary, both evaluation approaches indicate an appropriate accuracy of KONUX Switch in the light of its intended purpose to predict the need for maintenance of railway switches.

### 2.4.2 Robustness

The following paragraph demonstrates which robustness measures KONUX has taken to make the system resilient regards errors, faults or inconsistencies that may occur.

During the development of the model, KONUX ensures robustness of the systems at the data level by using real-world data and selecting an appropriate sample for training and testing of the model. Thus, unintended bias because of incomplete or inappropriate data can be prevented to a high extent. The imbalance of prediction classes was already discussed in the previous section. To counteract, an appropriate performance metric, i.e. the MCC, has been chosen and oversampling of underrepresented classes during training ensures data balancing in the development process of KONUX Switch. Additionally, the features the training of the model were selected carefully and with regard to existing research in the area as explained in 2.2. [8]

KONUX invested a lot of time in testing the KONUX Switch application with the range of all possible inputs sent by the switches including failures and abnormal status. Thus, the system knows to handle those inputs and displays an error status. For instance, a vertical displacement sent by a sensor that is outside of the reasonable range, that is calculated based on past data, will result in an error that informs the user about a possible failure of the components.

In addition, KONUX set up a backup plan for every application. A backup routine is triggered every 3 hours which ensures that the systems state can be recovered in case of a failure and thus all important data will be saved redundantly.

### 2.4.3 Cybersecurity

This paragraph demonstrates how KONUX reduces system vulnerabilities to the possible minimum and thus makes the system resilient regarding attempts by unauthorised third parties to alter the use of the system or the performance.

To develop appropriate technical solutions that ensure cybersecurity for KONUX Switch, the relevant circumstances and risks need to be identified. First of all, attackers could try to reduce the system's performance resulting in wrong maintenance predictions for switches. Thus, replacing switches that in fact have a longer RUL than calculated leads to high sunk costs affecting the profitability of KONUX Switch. Furthermore, missing the correct time for switch maintenance can result in failure with the consequence of delayed trains. This again leads to inefficiency in railway traffic which comes with high costs for railway operators. In extremely rare cases, the failure of a railway switch could induce damage on trains and potentially derailment of a train leading to physical harm of passengers. However, this extreme case is very unlikely and not necessarily a causal consequence of the KONUX Switch system. Second, unauthorized third parties could attack the system in a way that does not lead to incorrect predictions but to unavailability of the system. In this case, the same consequences as explained above apply. To summarize, both mentioned cases need to be prevented as they result in financial damage for our partners.

In the following, the most important attacks are described and the defense mechanisms by KONUX are explained. The first category of attacks is called data poisoning, which can be performed during the training stage. Attackers aim at tampering the features in the training dataset which might include updating/deleting data or injecting instances with incorrect labels into the training set. Consequently, the model fails to classify the data points correctly which lowers the performance for the intended use of the system. [9]

KONUX implemented various defense mechanisms against these type of attacks. First of all, our system performs data sanitation before training. During our development phase, the so called L2-Defense showed the best trade-off between security against the injection of incorrect data and limitation of the performance and generalization of the model. Consequently, they detect injected data by adversials using the L2-distance measure. Data instances, which are very distant from the center of the corresponding class, are discarded for the training of our model. Second, the team implemented to perform adversarial training in order to increase the strength of the model. This means that they added crafted data points into the training dataset which made the training process harder but at the same time the model learned to deal with incorrect examples. Third, they use the so-called transferability blocking defense mechanism. Concretely, they perform Null labeling which means that a null label is introduced into the training dataset and the model is trained to classify adversarial examples as instances of this null class. This results in discard of the malicious data. The advantage of the approach is that it does not reduce the performance of the model for normal data instances. [9]

The second type of attack that KONUX has secured the system towards are so called model poisoning attacks. Since poisoning of models can be rather seen as a traditional cyberattack, KONUX ensures secure hosting and disseminates pre-trained models in virtual repositories that guarantees integrity and prevents manipulation of the models.

Lastly, KONUX identified the need to protect the system against inference attacks that aim to gather information about the used train dataset and the origin of the data. In their opinion, leakage of the feature values can result in the publication of corporate secrets and financial or reputational damage of our partners. Thus, it should be strongly avoided. They use specific heuristics to evaluate how and how many entries need to be modified which have large correlations with the attribute values to any company that provided training data. [9]

## 2.5 EU Declaration of Conformity [Ozan Aydin]

**AI System Provider :** *KONUX GmbH Flößergasse 2 81369 München, Germany*
**AI System :** KONUX Switch

KONUX declares that the product KONUX Switch is in compliance with the following DIRECTIVES and STANDARDS, issued by EU:

- Directive 2006/42/EC of the European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (OJ L 157, 9.6.2006, p. 24)[as repealed by the Machinery Regulation];

- Regulation (EU) 2016/425 of the European Parliament and of the Council of 9 March 2016 on personal protective equipment and repealing Council Directive 89/686/EEC (OJ L 81, 31.3.2016, p. 51);
- Directive (EU) 2016/797 of the European Parliament and of the Council of 11 May 2016 on the interoperability of the rail system within the European Union (OJ L 138, 26.5.2016, p. 44).
- Directive 95/46/EC (General Data Protection Regulation)
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data
- Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts

This declaration of conformity is issued under the exclusive responsibility of the manufacturer, KONUX.
München, 12/06/2023
KONUX GmbH

## 2.6 AI System Performance in the Post-market Phase [Ozan Aydin]

In this chapter, as stated before, the post market monitoring system shall be set up such that it makes sure that the AI system will comply with articles 13 and 15.

### 2.6.1 Post-Market Monitoring Plan

In accordance with the Article 61(3) of the EU AI Act, the post-market monitoring plan shall be drafted and the post-market monitoring system shall be set according to the plan. The plan that KONUX Switch will follow consists of three major parts:

- Understanding the history and the overall status of the system before KONUX Switch is introduced
- Gathering relevant data throughout the AI systems lifetime
- Periodically analysing relevant data

In the following sections, the details for these steps will be revealed.

### 2.6.2 Understanding the History of the System

Unlike many AI systems KONUX Switch is used to solve tangible and calculable problems. The majority of the use cases revolve around end users realising that their current way of switch maintenance is inefficient, and they realise this because they have actual data at their disposal. Therefore, in order to figure out if KONUX Switch is staying accurate and robust in accordance with Article 15, the access to old data is crucial. This will provide a baseline for growth and will allow developers to analyze accuracy metrics properly.

### 2.6.3 Gathering Relevant Data

According to Article 61, the key part of any post-market monitoring plan is data collection. Apart from all the data KONUX Switch is collecting through the built-in sensors, for the post-market monitoring system KONUX Switch also aims to collect data regarding the performance of the system. For KONUX Switch, calculating the performance of the system after market release requires manual inspection. The service that KONUX Switch provides cannot be evaluated by any user feedback or user metadata, in order to understand if the data coming from the sensors are accurate the only way is to physically inspect the switches. So for this step, two main actions will be taken, backup of already incoming data and periodic manual inspection, which will be detailed in the upcoming two pharagraphs.

#### Manual Inspection

Manual inspection should be undertaken manually by KONUX Switch periodically. This allows developers to check if the predictions made with incoming data is accurate. For example, if the data is suggesting that a switch is broken or soon to be broken, a periodic manual inspection should verify if this is true, and this in turn will allow developers to measure if the system is in compliance with the necessary accuracy and robustness thresholds.

#### Backup of Already Existing Data

After manual inspection, if something does indeed go wrong, it is important for developers to pinpoint the exact error and display this to the end users, in accordance with the transparency regulations. In order to follow up on this goal, KONUX Switch will backup the incoming data from the sensors, so if there is a discrepancy between the actual manual results and the AI predictions, the exact data that the system picked up in that time shall be displayed to the end user, in order to simplify debug efforts. This will also allow developers to understand the root cause of the problem. For efficiency purposes, the stored data will only be stored between two manual inspections, and deleted afterwards to make way for the next wave of data.

### 2.6.4 Data Analysis and Report

After proceeding with data backup and manual inspection, the KONUX Switch team is determined to analyse this data and present the results to the end user in compliance with the transparency rules given in Article 13. Since the ground truth values are present because of the manual inspections undertaken by KONUX, the KONUX team can now begin their accuracy assessments with the actual data. How KONUX Switch does accuracy assessments is given in section Accuracy. KONUX will use the guidelines given to create an confusion matrix, and than compare the current accuracy results with the previous ones to figure out if the system is robust or if it is deteriorating. Afterwards, the findings need to be presented to the end user in a clear and concise manner. For this, a new section will be created in the dashboard, called the "Periodic System Checks". Here, the users will be able to find the current and the previous accuracy metrics and will be able to compare for themselves if the system is performing well. If indeed something has gone wrong, for example if the accuracy has dropped below a predetermined threshold, the system will show a warning. The end users will have the option to review the collected data themselves, but the main responsibility will be on the KONUX team, who will examine the data backups and figure out if the sensor is damaged or if anything else is performing poorly. Once the reason is clear, the KONUX Switch team will contact the end users and present them possible solutions.

# 3 Discussion and Ethical Analysis

In this last chapter, we will first present critical thoughts that emerged during the creation of the technical documentation. In the subsequent ethical analysis, we will discuss benefits and risks of KONUX Switch for different stakeholders and give an ethical assessment in the areas of data protection & privacy, discrimination, environment & sustainability, cybersecurity & physical safety and society. We also consider human rights and reflect on gaps in the EU AI Act, before we finally conclude this chapter.

## 3.1 Discussion of Technical Documentation

Due to the large scope, we will focus in the following on Art. 13 and Art. 15, which were mainly discussed in the technical documentation.

### 3.1.1 Discussion of Art. 13 [Janick Hofer]

As explained in detail in the previous chapter, Art. 13 deals with transparency and the provision of information. In the following, three aspects of this article will be examined in more detail.

With regard to transparency, it is worth to take another look at the definition of transparency. The current version of the EU AI Act defines transparency as interpretability and usability of the system's output. However, the definition of transparency is heavily discussed in the research area of Explainable AI. The research topic is trying to come up with a set of tools and frameworks to help understanding and interpreting predictions made by your machine learning models, contrasting the black box concept in machine learning. In our case, the EU AI Act's distinction between understandability and usability defines the problem well, such that, in the end, the user should be aware of the possibilities of the AI system.

KONUX seems to have already taken quite some steps into the direction of explainable algorithms by giving clear descriptions about the measures, as described in chapter 2. However, also KONUX itself will probably not know exactly how to calculate the predictions in the current development process, since deep learning is still mainly based on the black box algorithms. In this respect, it remains to be seen how KONUX will continue to improve its transparency efforts as research in Explainable AI research evolves. As Doshi-Velez et al. [10] also states, research will need to find this common taxonomy on which all machine learning applications will rely on. It will be interesting to observe whether the EU AI Act's definition of transparency will be this baseline.

Finally, we take a brief look at the second part of Art. 13, the provision of information. The amendments show that there is still disagreement among regulators about the scope of this information. In the concrete application of the law to Konux Switch, it can be stated that the scope of the data seems to be appropriate. Certainly, some information such as exact performance measures or exact data set splits for training will not be willingly released by the development team. Based on other precedent cases, it would be interesting to see what the scope of this specific information should be. The usefulness of human oversight measures also seems questionable, since the original idea behind predicitive maintenence is to relieve humans of manual oversight, and therefore renewed oversight would render the product obsolete again. In summary, some minor points can be criticized, but in general the scope of the required data seems reasonable.

### 3.1.2 Discussion of Art. 15 [Sophie Witt]

As described in section 2.4 Art. 15 of the EU AI Act states that high-risk AI systems should be designed to a achieve an appropriate level of accuracy, robustness and cybersecurity for their intended purpose.

In general, the term appropriate is very hard to define and does highly depend on the specific use case of an AI system. In particular, this is emphasized by the addition "in the light of the intended purpose" in Art. 15. Thus, the level of accuracy, for instance, should be appropriate for the purpose of the AI system. However, it seems that it is still up to the providers to decide what appropriate means for their specific high-risk AI system.

Regarding accuracy, there is a wide variety of different metrics that can be used. However, their meaningfulness depends on various factors which includes the data distribution, the AI task and the impact of errors. The latter one means that the implications of false positives and false negatives play a role in determining the meaningfulness of the accuracy. Depending on the specific application, the consequences of misclassification may vary in terms of financial, social, ethical, or legal impacts. In the case of KONUX, particularly financial and social impacts need to be considered. Additionally, related research areas provide guidance to select appropriate accuracy metrics and value ranges as outlined in 2.4.

Furthermore, the same concern applies to the appropriate level of robustness. In this case the assessment of appropriateness of the providers measures necessitates careful consideration of the system purpose, potential harms, complexity, environment, testing, validation, certification, and ongoing monitoring. Due to the specificity and the range of factors to take into account, it might not be feasible for providers to cover all aspects.

The same applies to the last obligation of Article 15 which is an appropriate level of cybersecurity. In this case, the main challenges for providers arise due to the implementation complexity, the rapidly evolving threat landscape, collaboration with third-party components and tradeoff between system performance (including usability) and security. For instance, the L2-defense implemented by KONUX (see 2.4.3) increases the security of the system but at the same can reduce the system's overall performance due to the discard of inputs that would have been valuable for the training of the model.

In summary, the obligations defined in Article 15 aim to ensure that high-risk AI systems perform consistently with an appropriate accuracy and that providers install the necessary robustness and cybersecurity measures. Consequently, this enhances the overall trustworthiness of these AI systems. However, due to the explained concerns, the required technical measures highly depend on the concrete use case of the application. Therefore, these obligations might be unachievable for some companies, especially smaller ones.

## 3.2 Ethical Analysis

### 3.2.1 Benefits and Risks of KONUX Switch for Stakeholders [Andreas Heckl]

Before identifying the benefits and risks of KONUX Switch, we have to consider who is potentially affected by them, i.e. we need to point out the stakeholders of a company that puts the AI system into use. These stakeholders are:

- employees of the company, e.g. workers who maintain switches
- owners or shareholders of the company (may also be the government, as in Germany or Austria)
- customers of the company, e.g. train passengers
- service providers or business-to-business (B2B) partners of the company, e.g. rail manufacturers

**Benefits** On their company website, KONUX states multiple advantages of KONUX Switch, for example prevention of defects, less corrective work and delay free switches [11]. Assuming their AI system works as well as desired, we identify the following benefits for the respective stakeholder groups:

- Employees that are concerned with the management of switch maintenance are supported by KONUX Switch and can make better decisions. This support allows them to allocate more time to other responsibilities, leading to increased efficiency. Likewise, inspectors and craftsmen can dedicate additional time to other elements of the railway infrastructure, aside from switches.
- After the initial investment for KONUX Switch is recouped, shareholders will ultimately benefit financially, as the system leads to cost savings, e.g. by less manual inspection.
- Customers are expected to face a more reliable service, as there should be fewer delays, resulting in a smoother travel experience.
- In our opinion, B2B partners such as manufacturers of switches and trains or software providers will not face any noteworthy benefits.

**Risks** Using an AI system for predictive maintenance in the rail sector not only provides benefits, but also poses risks to stakeholders. We identify the following risks for the respective stakeholder groups:

- Employees are at risk of losing their jobs, as the AI system could make switch management and maintenance more efficient and thus reduce the overall workload. This would affect employees who deal with the management of switch maintenance, inspectors and workers who actually carry out the maintenance work. Several studies have shown that job loss and unemployment and can lead to severe mental and physical health problems [12].
  Even if job loss is avoided and employees are assigned different responsibilities, the utilization of the AI system could still have a negative impact on the employees' self-esteem, since the AI now does the work faster, cheaper and more accurate than they did before. Further, employees with a comparatively negative attitude towards AI are less satisfied in their jobs and show less commitment to their work [13].
- As we have seen in the section on benefits, there are financial advantages for the shareholders in the case that the AI system works as desired. However, in the case that the system does not work as accurate as expected, there are also disadvantages: First, owners could suffer financial losses if the tool is inefficient or makes unreliable predictions that lead to higher maintenance costs. Second, there could be damage to the company's reputation if it relies too much on AI and no longer applies sufficient manual oversight. For example, the tool could fail to classify a switch that is in urgent need of maintenance as such. This may result in a severe accident and impair the public perception of the company.
- Similar to the shareholders, customers would only face risks in the case where the AI system is prone to errors or malfunctions that could for example lead to delays or longer travel times. In the worst case, a defect switch which is not recognized as such by KONUX Switch could be the cause of an accident which physically harms customers.
- In our opinion, most service providers and B2B partners will be unaffected by the use of KONUX Switch. However, manufacturers of switches and their spare parts could generate less revenue due to the more efficient management of the switches. Nonetheless, we consider this effect to be negligible and conclude that service providers are not exposed to any significant risks.

In our opinion, the benefits outweigh the risks. However, we suggest not to blindly rely on the tool, but recommend a combination of KONUX Switch and manual oversight to check whether the performance is accurate in practice.

### 3.2.2 Ethical Assessment [Andreas Heckl]

In this section, we will analyze the impact of the tool from an ethical perspective in the following five areas: data protection & privacy, discrimination, environment & sustainability, cybersecurity & physical safety and society. For each area, we will state our opinion on whether or not the usage of KONUX Switch is ethically reasonable.

**Data Protection & Privacy** The tool does not collect, process or analyze any person-specific data such as age, gender or religious beliefs. Therefore, we classify the use of KONUX Switch as ethically reasonable regarding data protection & privacy.

**Discrimination** As outlined in the previous paragraph, no personal data is utilized, which would be a necessary condition for the discrimination of persons or groups of persons. Therefore, we consider KONUX Switch to be ethically reasonable in terms of discrimination.

**Environment & Sustainability** Concerning environment & sustainability, opposing arguments can be made, depending on the performance of the tool in the real world.

First, we assume KONUX Switch works as well as desired. Among others, goals of KONUX Switch are to prolong asset lifetime and make maintenance more efficient. Given these goals can be reached, the tool would make rail transportation more attractive, e.g. due to fewer delays. In addition, cost savings could be passed on to customers or invested in the expansion of the rail network, also leading to a more attractive service for customers. Overall, more people would travel by train instead of by car or plane, resulting in less environmental impact, as train traveling has been shown to be the more environmental friendly option [14].

However, assuming KONUX Switch does not turn out to work as efficient as advertised, the above arguments can be reversed. For example, malfunctions of the tool could lead to more delays and cancellations of trains, making rail transportation overall less attractive for customers. This scenario could also lead to higher costs, which ultimately will be passed on to customers, also making rail transportation less attractive. For example, the tool could lead to unnecessary maintenance costs if it predicts a significant amount of false positives, i.e. classifying a switch to be in need of urgent maintenance, when in reality the switch does not need to be maintained in the near future.

In sum, regarding environment & sustainability, we classify KONUX Switch as ethically reasonable dependent on its performance in practice.

**Cybersecurity & Physical Safety** Like any software system that communicates over the Internet, KONUX Switch can be a target of cyberattacks. Hackers could, for example, falsify sensor data or prevent its transmission. An even worse threat would exist in the case where hackers could actually set the switches and thereby cause trains to collide, for example. However, KONUX Switch has no influence on actually setting the switches, but only evaluates their sensor data. So while there is a potential attack surface for cyberattacks, in our opinion, the type of possible attacks does not endanger the physical safety of train passengers directly.

Safety issues may not only arise from cyberattacks, but also from a low accuracy predictions. On a high level, the tool can make two types of wrong predictions: false-positives and false-negatives. We already discussed that the false-positives would lead to higher maintenance costs. A significant amount of false-negative predictions, i.e. classifying a switch not to be in need of maintenance when in reality it actually needs to be maintained soon, could also lead to problems, as switches may not work correctly any more. In the worst case, this could lead to a train accident and ultimately threaten the physical safety of passengers.

Accordingly, regarding cybersecurity & physical safety, we classify the use of KONUX Switch as ethically reasonable dependent on its performance.

**Society** We will present three arguments which indicate that KONUX Switch does not pose significant risks to society.

First, KONUX Switch is not a strong AI, but a weak one. This means that it cannot be applied to a diverse range of problem domains like computer vision, text generation and natural language processing at the same time, but rather is made for a very specific purpose, that is predictive maintenance of rail switches. In our opinion, using AI in such a narrow application domain will not have a large influence on society, as only a small amount of jobs are potentially affected. Also, from a financial perspective, we assume that the effect on the gross domestic product of a member state is negligible.

Second, KONUX Switch is not a generative AI that produces new content which could be potentially harmful to society like fake news or deep fake images. KONUX Switch only works with very specific data, that is sensor data of rail switches. The tool is mainly used for collecting and interpreting this type of data, ultimately leading to predictions and classifications.

Third, it is neither publicly available and nor for free, but can only be used by railway companies. This means that it is less exposed to the risk of being misused for unethical purposes, as is the case with AI systems that are freely available to anyone via the Internet.

From these three aspects, we conclude that KONUX Switch does not pose any significant risks to society as a whole, and thus classify the tool as ethically reasonable in this area.

We will provide an overall assessment of these five areas in the conclusion.

### 3.2.3 Human Rights Perspective [Suchithra]

Human rights are centered around protecting the fundamental rights of humans. They were created and adopted with support from the global populace in order to outline the rights and liberties that would enable every person to live in freedom and dignity. In general, it is evident that inclusion of AI anywhere significantly transforms the human experience. The EU AI Act is intended to address various AI system-related issues, including safeguarding human rights. It intends to provide a unified regulatory framework that oversees the creation, introduction, and use of AI technology within the European Union. The EU AI Act specifies certain demands and duties for high-risk AI systems, including transparency, accountability, and human supervision, which may help protect human rights and resolve ethical issues.

It is clear that applying AI to the rail sector might improve both efficiency and quality. But it is also crucial to consider this from a human rights standpoint. Article 2 of the Universal

Declaration of Human Rights ("UDHR") [15] and provisions of the International Covenant on Civil and Political Rights ("ICCPR") especially the ones laid down under Article 2 and Article 19(2) [16] state that each person has a right to all freedoms and rights without restriction. The fundamental human rights granted under the UDHR and ICCPR state that there should not be any discrimination on the basis of race, color, sex, language, religion, political or another opinion, national or social origin, property, birth, or other status. The United Nations Human Right provisions also recognize the "right to work, which includes the right of everyone to the opportunity to gain a person's living by work which he freely chooses or accepts and will take appropriate steps to safeguard this right. "

KONUX Switch does not invade these human rights and the ethical guidelines[17] outlined per se. Most of the KONUX Switch's techniques are widely employed and deployed in diverse contexts throughout the globe in the transportation sector by other providers who supply similar AI tools. The main difference is that the KONUX Switch has improved the current experience of public transportation users. KONUX Switch providers have tried to implement necessary checks and balances to guarantee adherence to openness, privacy, and human rights duties.

Understanding the advantages and hazards of AI is crucial. In the words of the United Nations Human Rights Chief, "Artificial Intelligence can help society solve some of its biggest problems. However, if applied without consideration to human rights, AI technology may have dire consequences"[18].

### 3.2.4 Gaps and Aspects Currently Not Covered Under the EU AI Act [Suchithra]

Evaluation of KONUX Switch alongside the EU AI Act has given us a broader perspective of the EU AI Act and also the need for the improvisation of the EU AI Act to ensure a mid-path for AI providers and AI users. Here are a few gaps that may be considered to improve the current Act and guarantee greater adherence to ethical standards. It is vital to ensure equity and correct biases in AI systems for their sustainability and societal adoption.

The EU AI Act strongly emphasizes openness, but it could be more precise about what constitutes explainability and interpretability of AI choices. Fixation of accountability and societal confidence in AI systems depend on our ability to understand how they get to their decisions. To address possible problems brought on by AI technology, it is vital to establish clear lines of accountability and culpability. It is also necessary to consider AI's broader social and economic effects, such as job loss, economic inequality, and societal instability, which may require extra consideration and mitigating measures.

From our analysis of KONUX Switch, we have concluded that specific measures should be considered to solve the issues and assure adherence to ethical norms, notably in the context of predictive analysis AI tools, to govern AI in public transportation, including trains, and enhance the present Act. Here are some possible additions that may be made to specific Articles of the EU AI Act:

- Explainability and Interpretability (Article 22)

There is a requirement to outline the conditions for the interpretability and explicability of AI systems employed in public transportation, mainly tools for predictive analysis. Suppliers must comprehensively explain how AI-driven predictions and choices are generated for stakeholders to comprehend the underlying logic and algorithms.

- Accountability and transparency (Article 24)

To ensure that AI systems used in public transportation, mainly tools for predictive analysis, are subject to clear lines of responsibility, the regulations relating to accountability and transparency should be strengthened. To ensure the ethical use of AI technology, it is necessary to clearly define the roles and responsibilities of stakeholders, including operators, developers, and regulators.

- Impact Assessment (Article 28)

Introduce a need for impact analyses unique to AI systems used in public transportation, especially those that use predictive analysis. These evaluations should consider the possible social and economic repercussions of the sector's adoption of AI, such as job loss, economic inequity, and societal instability. To deal with any dangers, mitigating measures should be found and implemented.

- Training Data Bias (Article 10)

Strengthen the clauses addressing bias in training data for AI systems, such as predictive analysis tools, in public transportation. This might include rules and procedures for assuring varied and representative datasets to eliminate biases that can result in discriminatory results and promote justice and equality in decision-making.

- Human Oversight and Intervention (Article 16)

In the context of predictive analysis, emphasize the value of human monitoring and involvement in AI systems employed in public transportation. To achieve this, it is essential to specify the responsibilities of human operators, define rules for their participation, and ensure AI systems are built in a way that allows for appropriate human control and intervention.

A legislative framework that meets AI difficulties and public transportation needs may be created, notably in predictive analysis tools, by including these particular inputs into the EU AI Act. This would promote moral behavior and maintain justice, accountability, and public trust essential for developing AI in critical infrastructure areas like public transportation and railways.

Despite this, it is understandable that no legislative system can completely handle all ethical problems or all potential gaps. Since AI is a technology that is constantly developing, new ethical issues can arise in the future. Although the EU AI Act offers a robust framework for regulating AI systems, it is critical to routinely review and amend the laws to ensure they remain valuable and adaptable to the rapidly changing field of AI technology. Additionally, continued partnerships among legislators, academics, business professionals, and members of civil society may aid in identifying and resolving any possible gaps or newly developing ethical challenges that are not sufficiently addressed by current legislation.

## 3.3 Conclusion [Andreas Heckl]

In this chapter, we first covered Art. 13, which deals with transparency and the provision of information. We outlined that there is a technically inherent conflict between black box AI methods and transparency. It remains to be seen how technologies will develop further and also how research in the field of Explainable AI evolves in the years to come. From our point of view, the approach KONUX takes to ensure transparency and the provision of information as required by the EU AI Act is appropriate. Second we discussed accuracy, robustness and cybersecurity, all of which are covered in Art. 15. We stated our concerns regarding these three properties and concluded that they might be unachievable for some companies.

Based on our analysis of benefits and risks of KONUX Switch, we conclude that the use of AI in the rail sector is worth a try. Naturally, the utility of such a tool ultimately depends on its performance in the real world. It is worth mentioning that overall, AI technologies are still in their early stages, and we anticipate significant advancements in the years ahead.

From an ethical perspective, we outlined that we have no concerns about KONUX Switch in the areas of data protection & privacy, discrimination and society. Regarding environment & sustainability as well as cybersecurity & physical safety, we were not able to definitely classify the system, because both advantages and disadvantages may exist. Consequently, we did not find any area in which we would rate the tool as ethically unreasonable. In sum, we consider the possible disadvantages to be tolerable and believe that, with due care, e.g. by combining KONUX Switch with manual inspection, no major threats to stakeholders or society will arise from this tool.
Moreover, we do not see any violations of human rights that originate from this AI system.

Finally, it remains to be seen how effective and accurate KONUX Switch can be applied in the rail sector in the coming years, and to which degree the EU AI Act will have an impact on it. What is certain, however, is that technology and regulation are in a constant state of change and must continuously adapt to each other.

# Bibliography


[1] https://www.konux.com/konux-final-close-brings-series-a-to-16m/. Accessed: 2023-06-16.

[2] KONUX GmbH. *KONUX Switch - Demo Video*. Mar. 2022. (Visited on 06/02/2023).

[3] T. Böhm. "Remaining Useful Life Prediction for Railway Switch Engines Using Artificial Neural Networks and Support Vector Machines". In: *International Journal of Prognostics and Health Management* 8 (Dec. 2017).

[4] B. Kyurkchieva. *Spring '22 Release: Introducing Our New and Improved User Interface*. https://www.konux.com/spring-22-release-introducing-the-new-and-improved-user-interface/. Mar. 2022. (Visited on 06/02/2023).

[5] B. Kyurkchieva. *Performance at Scale*. https://www.konux.com/de/next-generation-iot-release/. Sept. 2019. (Visited on 06/03/2023).

[6] *Digitalisierung: Konux erhält Nachweisführung der DB*. https://www.eurailpress.de/nachrichten/unterneh maerkte/detail/news/digitalisierung-konux-erhaelt-nachweisfuehrung-der-db.html. (Visited on 06/02/2023).

[7] G. Jurman, S. Riccadonna, and C. Furlanello. "A Comparison of MCC and CEN Error Measures in Multi-Class Prediction". In: *PLoS ONE* 7.8 (Aug. 2012). Ed. by G. Biondi-Zoccai, pp. 1–5. DOI: 10.1371/journal.pone.0041882. URL: https://doi.org/10.1371%2Fjournal.pone.0041882.

[8] *Practical Ways To Improve The Robustness of Machine Learning Models*. https://www.periculum.io/post/machine-learning-models. Accessed: 2023-06-02.

[9] R. S. Sangwan, Y. Badr, and S. M. Srinivasan. "Cybersecurity for AI Systems: A Survey". In: *Journal of Cybersecurity and Privacy* 3.2 (2023), pp. 166–190. ISSN: 2624-800X. DOI: 10.3390/jcp3020010. URL: https://www.mdpi.com/2624-800X/3/2/10.

[10] F. Doshi-Velez and B. Kim. *Towards A Rigorous Science of Interpretable Machine Learning*. Mar. 2017. DOI: 10.48550/arXiv.1702.08608. arXiv: 1702.08608 [cs, stat]. (Visited on 06/08/2023).

[11] https://www.konux.com/. Accessed: 2023-06-16.

[12] D. Dooley, J. Fielding, and L. Levi. "Health and unemployment". In: *Annual review of public health* 17.1 (1996), pp. 449–465.

[13] D. Brougham and J. Haar. "Smart technology, artificial intelligence, robotics, and algorithms (STARA): Employees' perceptions of our future workplace". In: *Journal of Management & Organization* 24.2 (2018), pp. 239–257.

[14] S. Boschmans, I. Mayeres, and B. v. Zeebroeck. "Transport and Environment Report 2020 Train or Plane?" In: (2021).

[15] `https://www.un.org/en/about-us/universal-declaration-of-human-rights`. Accessed: 2023-06-16.

[16] `https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights`. Accessed: 2023-06-16.

[17] `https://unesdoc.unesco.org/ark:/48223/pf0000381137`. Accessed: 2023-06-16.

[18] `https://news.un.org/en/story/2021/09/1099972`. Accessed: 2023-06-16.