Managing AI use in telecom infrastructures

Advice to the supervisory body on establishing risk-based AI supervision

Commissioned by:
Dutch Radiocommunications Agency

Project:
2019.166

Publication no.:
2019.166.2004 v1.2.6

Date:
Utrecht, 17 June 2020

Authors:
ir. Tommy van der Vorst
ir. Nick Jelicic
ir. Jan van Rees
prof. dr. ir. ing. Rudi Bekkers
ir. ing. Reg Brennenraedts MBA
Roma Bakhyshov MSc

Management summary

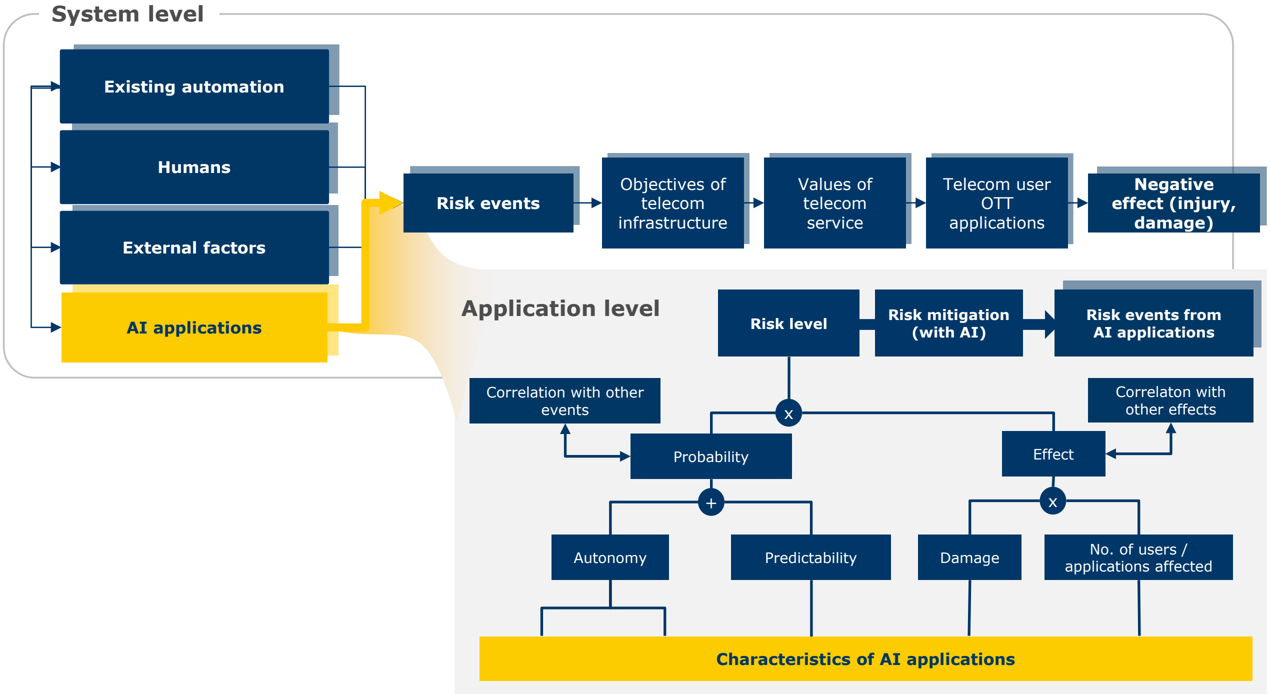
Telecom infrastructures are of vital importance to society. More and more applications depend on well-functioning, reliable and always available telecom services. The Dutch Radiocommunications Agency (Agentschap Telecom) oversees these in the Netherlands. The emergence of Artificial Intelligence (AI) applications has not only fundamentally changed the nature of the telecom sector but also the risks. In order to safeguard the proper functioning of the telecom infrastructure, adjustments are required in the relevant knowledge and in the approach to supervision policy. This report provides insight into how AI applications impact and endanger the telecom landscape and suggests how the Radiocommunications Agency can continue to maintain society's trust in the telecom infrastructure. The three methods used in this study are: literature research, interviews and sessions with experts.

Responses to the research questions

What are the current and future risks of applying AI in the telecom sector?

Bearing in mind recent developments, a relevant description of AI is: using algorithms based on deep learning, and learning assisted by big data, to automate tasks that could formerly only be undertaken (properly) by humans. AI is expected to play an increasingly central role in telecom networks.

AI applications have *specific characteristics* that can pose risks for telecom infrastructures (the application level). Various AI applications interact with each other, with people, 'normal' automation and possibly the outside world. It is therefore important to assess how AI is applied in the telecom sector at a *systemic level;* that is to say looking at the effects and risks for the entire chain rather than AI applications in isolation. In particular, they include the ultimate use of these applications based on telecom infrastructures.


System level
Existing automation
Humans
External factors
AI applications
Risk events
Objectives of telecom infrastructure
Values of telecom service
Telecom user OTT applications
Negative effect (injury, damage)
Application level
Risk level
Risk mitigation (with AI)
Risk events from AI applications
Correlation with other events
Correlaton with other effects
x
Probability
Effect
+
x
Autonomy
Predictability
Damage
No. of users / applications affected
Characteristics of AI applications

At application level, the extent of autonomous learning and operations, as well as the unpredictability, action framework and sphere of influence of AI applications determine the probability and impact of the additional risks. On top of the entire lifecycle of an AI application, including planning, data collection, training, testing, validation and operations, are notwithstanding the conventional risks relating to information security.

Although AI applications introduce new (types of) risks, ultimately they can add specific value to the process of mitigating risk.

How can the Dutch Radiocommunications Agency as supervisory and implementing organization mitigate these risks?

We recommend starting with tools at a systemic level. The Dutch Radiocommunications Agency could mitigate the risks of AI applications in the telecom sector by providing information and raising awareness, stipulating transparency, facilitating risk analysis and mitigation, as well as developing criteria and setting process requirements. There are specific tools for dealing with certain AI risk factors. At application level, more specific tools could be implemented: certification, auditing and maintaining particular types or aspects of AI could play a role. In a wider sense, there should be a social debate about the desirable level of telecom infrastructures' provision.

What does the use of AI look like now and in the coming five years for the telecom sector and other sectors that use digital connectivity?

Most of the current AI applications focus on improving specific parameters. These are strictly defined applications such as optimising the parameters of a radio signal, power management or routing traffic through a network.

Looking at the coming five years, we see AI applications becoming more and more advanced. Several suppliers of telecom equipment share the view that AI will control the majority of functions in telecom networks. Although it is questionable whether this will be implemented (entirely) in five years, their vision is definitely one we can expect.

How do we weigh up the risks to the various interrelated aspects in a risk model for digital connectivity?

Certain characteristics of AI applications can pose additional risks for telecom infrastructures. These characteristics relate to the following aspects of AI:

- **The extent of autonomous learning and implementation of AI.** If this extent is considerable, the likelihood of risk events increases. A significant parameter is whether the AI application is controlled by people or by rules.
- **The extent of the AI application's predictability.** If the models are non-deterministic or highly non-linear, it is more complicated to assess whether an application will work well in all situations. One influential factor is the type of data used and if it can be manipulated.
- **The AI application's action framework.** If the AI application has a highly limited effect on telecom infrastructures, this restricts the impact of a risk event. An application with a wide operating framework has a potentially greater impact.
- **The AI application's sphere of influence**. An application operating at a central level and controlling a telecom infrastructure is more prone to risk than an application that optimises a specific parameter at a low level.

The diagram below is an overview of the relevant aspects and their weighting. The scores can be combined under the "application level".

Autonomy			Predictability			Damage		Scope		
Learnin g	Validation	Action	Transparen cy	Deterministic & linear	I/O	Action framework	Use	Scope	Redundancy & variation	Score in risk model
Continu lerend	Unvalidat ed	Closed loop	Intranspa-rant, 'black box'	Non deterministic & non linear	Unconstrained/ untrusted	Broad action framework	Closed loop	Central	Non-redundant element	10
●	Model not seen nor tested	AI in closed loop	Model not seen nor certified	Stochastic AI-algorithms	Use unlimited data set	Network set-up	AI in closed loop	Network orchestrator	●	
●		●	●	●	●	●	●	●	●	
●	●	Constrain ed closed loop	High number of para-meters	Sequence sensitive (RNNs)	Use 3rd party data	●	Constrain ed closed loop	Edge	Redundant element	
Online lerend	●	●		●	●	Control network element	●	●	●	
●	●	●	●	Non-linear elements	●	Control traffic	●	Base station POP, MDF	●	
●	A few scenarios tested	Human in closed loop	●	●	Use own network data	●	Human in closed loop	●	●	
●		●	Training data unknown	●	●	●	●	CPE	●	
Een-malig getraind	All input combinati ons tested	AI in open loop	Certified	Linear regression	Only generated data	Optimali-sation of parameter	AI in open loop	Handset, terminal	Redundant varied element	
Offline lerend	Validated	Open loop	Explicable 'white box'	Deterministic & lineair	Constrained & trusted	Limited action framework	Open loop	Decentralise d, isolated	Double redundant, varied element	1

Characteristics of AI applications

Considering the risks of AI applications in isolation paints a limited picture of the societal risks (as well as advantages) of implementing AI in telecom infrastructures. At the systemic level, the following factors affect risks:

- **Interaction between AI applications and other systems.**
- **Replacing humans with AI.** Having people carry out tasks involves risks, and these can be higher or lower with an AI application. This study does not chart the risks involved with human activities in telecom infrastructures. The model we present can be used to assess the risks of substituting with AI in order to inform the decision whether or not to implement a human-replacement AI application.
- **Implement AI applications to mitigate risk.** At a systemic level, AI applications can contribute to lowering the level of risk, for example through faster detection of problems or attacks, and by helping to find causes and solutions.
- **Cyber (in)security of AI applications.** AI applications are of course also subject to cyber threats and associated security risks. These risks may increase, because training AI applications involves bringing together large amounts of (sometimes sensitive) data.

Table of Contents

Management summary 3

Responses to the research questions 3

1 Introduction 9

1.1 Background 9

1.2 Research questions 9

1.3 Approach 9

1.4 Reader guide 10

2 The rise of AI in telecoms 11

2.1 What is AI? 11

2.2 Risks when using AI 15

2.3 Current AI applications in telecom 18

2.4 Future applications of AI in telecom 21

3 Risks of using AI in telecom infrastructures 25

3.1 Systemic level 25

3.2 Application level 31

3.3 Determining risk 42

4 Role of the Dutch Radiocommunications Agency 45

4.1 Action framework 45

4.2 Assessing risks 46

4.3 Tools 49

5 Conclusions 51

5.1 Response to the main question 51

5.2 Responses to the subquestions 51

6 References 55

Annex 1 List of interview participants 57

Please cite as: Dialogic, van der Vorst, T., Jelicic, N., et al. (2020). *Managing AI use in telecom infrastructures. Advice to the supervisory body on establishing risk-based AI supervision.* Dutch Radiocommunications Agency, Groningen, the Netherlands.

Originally appeared as Dialogic, van der Vorst, T., Jelicic, N., et al. (2020). *Gebruik van en toezicht op AI-toepassingen in telecominfrastructuren. Advies aan de toezichthouder over inrichting van risico-gebaseerd AI-toezicht.* Agentschap Telecom, Groningen. Translation to English by Val Kidd.

1 Introduction

1.1 Background

Artificial Intelligence (AI) has been clearly advancing since around 2010. Although the concept was known and applied from the 1940s already, advances in computing power, storage capacity and telecommunications in recent decades have paved the way for novel potential uses. New algorithms based on deep learning use data to learn and perform tasks that formerly could only be done by humans. As such automated systems can process an incomparably greater amount of data than human beings, they open up new opportunities but also new problems: the systems are more complicated to analyse and for people to understand. Applying AI thus poses several social and ethical issues.

AI is already applied in a wide range of sectors including telecoms. These telecom infrastructures are vitally important for society. More and more applications depend on efficient, reliable and always available telecom services. In this study we consider the potential risks associated with using AI in telecom infrastructures.

1.2 Research questions

This study answers the following question:

What risks are involved in the current and future application of AI in the telecom sector, and how can the Dutch Radiocommunications Agency mitigate these risks?

We also attempt to answer the following sub-questions:

1. What does the application of AI look like now in the telecom sector and other sectors that make use of digital connectivity?
2. What developments are envisioned in the coming five years[1] for applying AI in the provisioning and use of digital connectivity?
3. What are the risks regarding availability, authenticity, integrity, trust, transparency and predictability in the various sectors as a result of the current and future use of AI? How do we weigh up the risks to the various interrelated aspects in a risk model for digital connectivity?
4. How can the Dutch Radiocommunications Agency as supervisory body and implementing organization mitigate these risks?

1.3 Approach

Three methods were used to answer the study questions: a literature review, interviews and validation sessions with experts. The literature was especially helpful for questions 2 and 3. Reviewing the literature can show what kind of AI applications are possible, although these are not necessarily used in the Dutch telecom sector.

To explore future AI developments, this study looked into scientific research on AI in the telecom sector. This helped to determine which lines of research could later be translated to

[1] A five-year time horizon seems rather short for an investigation like this. Nevertheless, we note that developments in the AI field, within and beyond telecoms, are in full swing. Looking *back* to five years ago, it would have been quite a challenge at that time to predict the status of AI today.

specific research by providers. We also looked at providers' white papers and road maps to identify future innovations. Similarly, we identified mitigation strategies for AI risks.

The interviews provided additional information relating to all the study questions. By interviewing telecom operators, it was easier for us to find out which telecom networks in the Netherlands are actually applying AI now and will do so in the near future. We spoke with suppliers of network equipment to hear more about research and development trends. In this way we were able to chart developments as well as supply and demand.

1.4 Reader guide

We start by describing in section 2 the AI developments we are currently witnessing in the telecom sector. We discuss what AI means for the sector at this time and for the coming five years: what AI applications do we anticipate in the telecom sector, what opportunities will they present and what risks generally go hand in hand with using AI? In section 3 we present a model for assessing the risks of AI applications specifically in telecom infrastructures. Section 4 looks at the role the Dutch Radiocommunications Agency can fulfil in mitigating the risks of AI applications in telecom infrastructures. Finally, we answer the research questions in section 5.

2 The rise of AI in telecoms

We start this section by defining Artificial Intelligence: what is it and why is it relevant in the context of risk? We then look at the current and future application of AI in telecom infrastructures.

2.1 What is AI?

The term AI has been around since the 1940s. As it has taken on another meaning over the years, for this study we define AI as follows:

Artificial Intelligence or AI is the use of algorithms based on deep learning, taught using large amounts of data, to automate tasks that could previously only be performed (properly) by humans, or to a limited extent by traditional automation.

We have chosen a working definition tailored to the framework of this research, not a normative definition. If we take a broader look at human tasks now being done by machines, many things such as a pocket calculator come under the category of “artificial intelligence”. However, the research question relates specifically to a (perceived) new “wave” of AI applications in the telecom sector. These applications form a new generation of AI, typically using deep learning algorithms, huge computing power, and large amounts of data. In light of this combination of ingredients, it is essential we consider the risks. As we will explain, they lead to systems that are even more complicated to understand and to control.[2] We start by underpinning this definition in the historical context.

2.1.1 History of AI

The history of AI goes back further than many would suspect: the dream of automating human behaviour and reasoning can be traced to ancient times. Greek mythology describes the giant bronze automat Talos, ingeniously created to protect the island of Crete from pirates. We find AI ideas in the medieval legends about the Golem of Prague protecting Jews. [1] Modern concepts of AI stem from the rise of the computer, with key figures like Turing, Walters and Minsky.

The term “artificial intelligence” was introduced in 1956 by scientist John McCarthy. [2] Since then, the scientific field has seen a number of cycles: highs with a great deal of AI hype, followed by lows, with disappointment and criticism (the so-called AI winter). So far there have been three major revivals and two relapses. The first revival of the term took place around the 1950s and 1960s, driven by pioneers at MIT and Stanford. The 1970s, however, saw severe cuts in research budgets. It appeared that AI could not for example translate Russian texts into English – an application in demand and expected AI could fulfil at that time. The classic example is how AI translated the Russian equivalent expression “*the spirit is willing, but the body is weak*” as “*the vodka is good, but the meat is spoiled”.* [3]

In the 1980s, Japan advanced its industry by firmly committing to AI. The United States and the United Kingdom soon followed suit. The emphasis in this period was on expert systems that emulated specific human actions. The intelligence was still programmed entirely by hand

[2] It is conceivable that our study findings also apply to AI uses that fall (just) outside this definition. For example, some of the risks found in Chapter 3 apply to self-learning systems, even though these are not based on deep learning.

and the system could not learn new tasks without humans programming the new rules. These systems are best described as a pre-programmed "decision tree" systematically running software. Figure 2 is an example of such a decision tree (an expert system would of course have a much larger "tree" to facilitate more complex decisions).

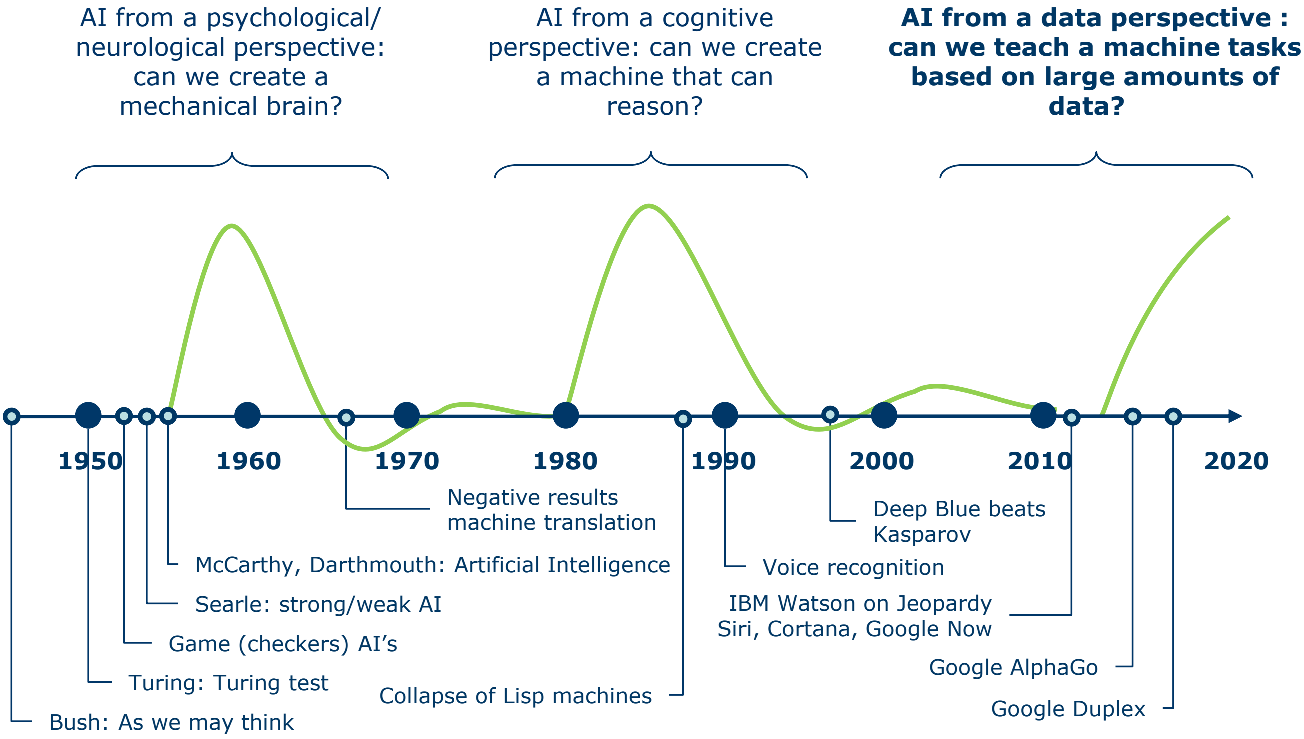


Figure 1 Overview of the history of Artificial Intelligence showing its three distinctive "waves"

Expert systems are usually complicated to devise. They require a combination of domain knowledge and programming know-how in order to develop software powerful enough to make "human" decisions. Despite the considerable hype surrounding expert systems, the high expectations have not been met. In the 1990s, the focus on AI dwindled. Nevertheless, expert systems have been adopted successfully to automate decisions, also in the telecom sector. [4]

Around 2011, AI experienced a revival thanks to the input of researchers like Andrew Ng, [5] Geoffrey Hinton [6] and Yann LeCun. [7] They developed *deep learning* – techniques that can make headway with the intelligence of algorithms. AI applications that up until then were deemed impossible, suddenly became achievable. One example is *AlphaGo,* developed by Google, which in 2016 beat world champion Lee Sedol in the game Go, even though up till then people assumed Go could only be played at a high level with human intelligence (and intuition). The game Go has 10,172 potential board positions [8] and 361 potential moves in each turn – many times more than the number of board positions and moves in the game of chess. This number of moves is too high for determining an optimal strategy with traditional methods, for example with a *minimax tree search.*[3]

[3] A minimax is a decision rule stating the best choice is the one that prevents a worst-case scenario. In a chess computer, this means that the best move is the one with the least chance of losing a piece. The effectiveness of a minimax tree search depends on the number of future steps observed. Although minimax is a simple rule, IBM applied it successfully to defeat Gary Kasparov. [47]

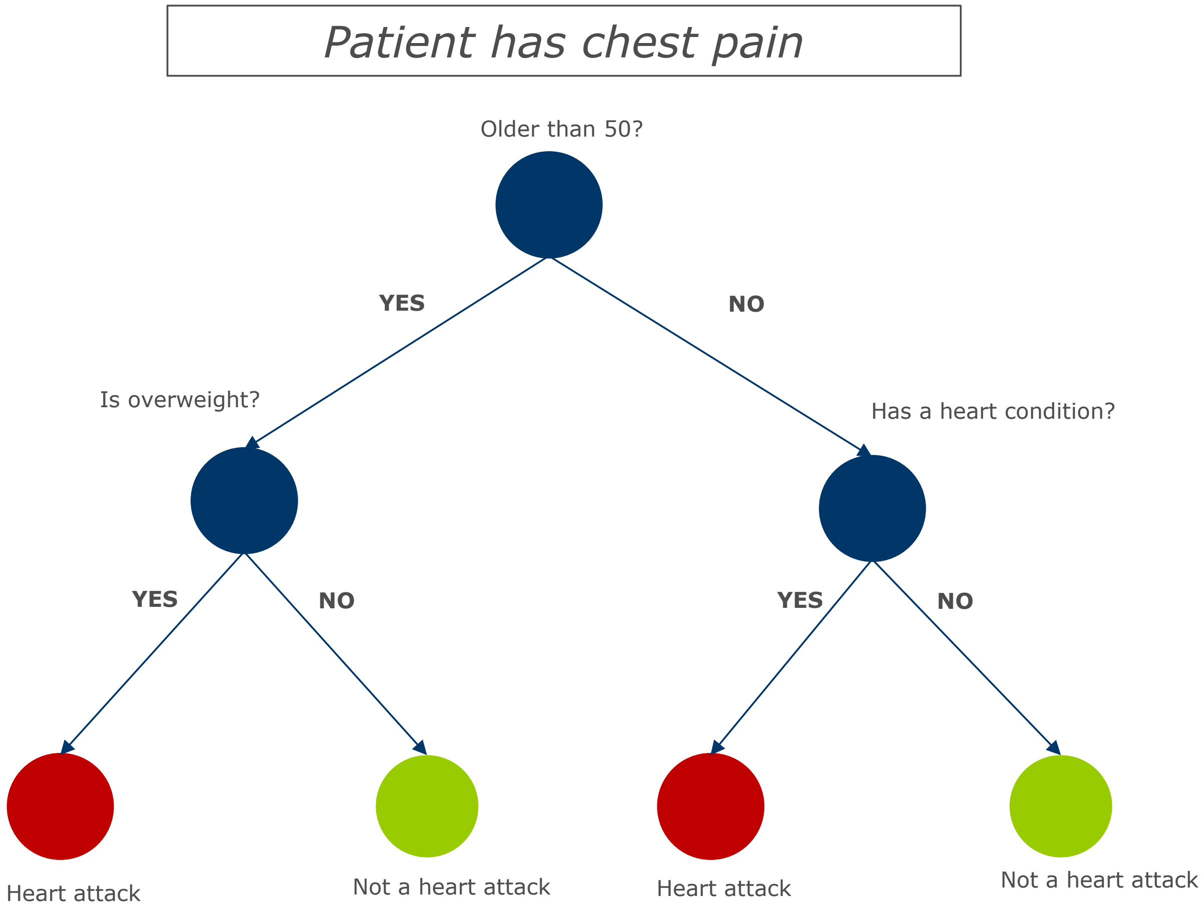


Figure 2 Applying a "decision tree" in First Aid to help determine: has the patient with chest pain suffered a heart attack?

What is machine learning?

AI can be achieved in different ways. At the time of expert systems, a system's intelligence was programmed manually. The system designers had to specify all the potential AI actions themselves. Consequently, the intelligence was quite limited: the system could not deal with a situation if the designer had not provided a rule for it. Nowadays, AI systems are not driven by rules but by data: the AI systems learn the rules themselves from the data. The underlying algorithms of these self-learning systems are typically called *machine learning algorithms*.

A specific category of machine learning is *deep learning*. This is an iterative search for mapping between the input and output of a model in the form of a series of mathematical transformations. These transformations are inspired by the way our human brain works: a so-called "neural network". In our brains, our senses stimulate brain cells – the "neurons". Depending on the stimulus, these neurons may or may not send a signal to other neurons. Hundreds of billions of neurons combined lead to intelligent behaviour.

A neural network consists of several layers: an input layer, multiple hidden layers and an output layer. The *hidden layers* extract properties from the data based on input. The properties of the final hidden layer are ultimately used to construct the output. In all the layers of the network, each neuron multiplies the input from the previous layer, multiplies it by a weight, adds up all these multiplied inputs and applies non-linearity. Figure 3 is a schematic representation of a neural network with 3 hidden layers.

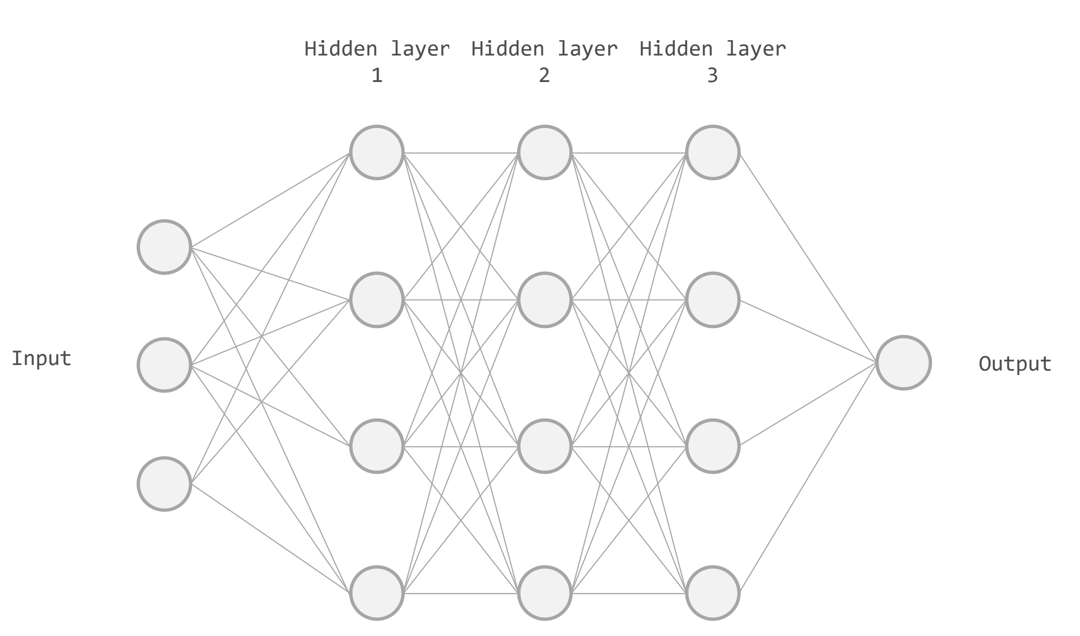


Figure 3 Schematic representation of a neural network

A rule of thumb for machine learning is that the model's performance improves as soon as more data becomes available for training the model. Advances in the digitization of processes have generated a lot of data that can be used for this purpose. Combined with ever-increasing computing power and academic breakthroughs, [8] [9] developments in machine learning are gaining more and more ground.

Deep learning has many applications that come close to or surpass human performance, certainly in calculating power and speed. In the medical world, deep learning is used to diagnose skin cancer and classify CT scans. In self-driving cars, deep learning determines the direction and speed of the car based on cameras. In the fight against fake news, Facebook among others uses deep learning to assess the authenticity of texts.

What will AI look like in fifty to a hundred years? In *science fiction,* we see futuristic images of machines acting as if they are human beings, and in many cases even have superior intelligence. A level of AI with an intelligence similar to that of humans is called *artificial general intelligence* (AGI); if the AI is smarter than humans, then we speak of *artificial super intelligence* (ASI). Some futurologists believe that AGI and in particular ASI could be man's final invention. AGI or ASI could even be a threat to mankind, since we as humans can no longer keep up with the superior intelligence of such AI. [10] Obviously ethical aspects could come under pressure.

Opinions are divided on whether (and when) this will ever happen. Many experts agree that deep learning in its current form is probably not suitable for producing human intelligence. [11] [12]

2.1.2 *Added value of AI*

AI generally provides a number of advantages and opportunities:

- **AI can make faster and sometimes better decisions than people.** Whereas a person sometimes needs a few seconds to minutes (depending on the amount of information to be processed) to make a decision, a machine learning model can often process thousands of data items in a fraction of a second. For example, fraud detection AI can monitor thousands of credit card transactions in *real time* and block potentially fraudulent transactions.

- **With AI, (scarce) expert knowledge can be used more efficiently.** People often undergo training for several years before entering the labour market. Even a few years later, a person will not yet be at the top of their game. Expert knowledge is therefore scarce and difficult to scale up. With machine learning, an expert's knowledge can be distilled into a model and thus this knowledge can be applied more widely. For example, Google has developed an AI system that can identify tumours from CT scans just as well as an experienced radiation oncologist. [13] Such application of AI can provide scarce expert knowledge in certain scenarios and thereby free up the "real" experts to do other work more efficiently.

- **AI is good at repetitive tasks.** People often perceive such tasks as not giving satisfaction. However, if tasks are well framed, AI is ideally suited to take them on. AI does not have to sleep, rest, or take breaks because it will not get bored or tired. Repetitive tasks also lend themselves well to AI because (if the task is currently performed by humans) there is probably sufficient data available to train the AI.

2.2 Risks when using AI

Compared to traditional automation and (non-AI) algorithms, AI-based systems have a number of unique, new characteristics. If these are not sufficiently taken into account, generically speaking, they entail risks. These risks are inherent in the underlying effects of AI and exist regardless of the application domain. In this section we describe these characteristics, how they translate to risks, and provide specific examples in the application domain of telecom.

2.2.1 Lack of accountability leads to further uncertainty in decision making

An AI system generally translates a set of "input" variables into a certain outcome ("output"). In deep learning-based AI systems, it is not evident how a particular outcome is achieved. Consequently, it is not always clear how a decision came about, it is complicated to verify certain actions, and errors can creep into the system undetected.[4] The authenticity of decisions made on the basis of the AI application may also be called into question.

[4] Incidentally, the relationship between the number of neurons and intelligence is a topic for discussion. For example, it is not the number, but precisely the extent of Interconnection between neurons that determines the degree of intelligence. Artificial intelligence may lead to new insights here once the hypotheses for artificial intelligence have been tested. Note that many telecom infrastructures without AI are already complex systems, and we might question to what extent inexplicable (or incomplete) choices are already being made.

Example from the telecom sector: Covariance shift

An *anomaly detection* system can detect and block suspicious traffic in a network. Such a system is trained to recognize abnormal patterns. Imagine that when a new browser is launched in the market, it applies a new protocol to better streamline HTTPS requests. As a result, a *covariance shift* takes place.[5] In the past, during training, the anomaly detection system has not seen any traffic passing through this browser. As a result, this browser's traffic is noted as “different” and blocked in the network.

2.2.2 Unpredictability leads to a lack of trust

In order to entrust AI systems with certain decisions, it is often important that the AI system's behaviour and outcomes are predictable. [14] There are at least two reasons why an AI system (despite it being a piece of software and its actions fully traceable), is capable of exhibiting unpredictable behaviour:

- **The algorithm is too complex for a human being to understand its behaviour.** Although conceptually we can imagine that an AI model consists of a large number of layers with functions in between, a model's precise behaviour is hardly or even no longer understandable or traceable in very large models. Currently we see AI models in use with hundreds of millions of parameters. [15]

- **The AI algorithm is not deterministic.** Many “ordinary” algorithms are deterministic: if you run them twice with the same input parameters, you get the same result both times. However, this is not the case with all algorithms: some use randomness, such as Bayesian methods. [16] In these models, the parameters are not fixed values, but distributions of opportunities. When a prediction is made with these models, random samples are taken from the distributions, and the chances are very slim that the same sample is drawn twice. AI algorithms often belong to this latter category, or have similar, non-deterministic characteristics. If the algorithm, and therefore the result, is not deterministic, the application is more difficult to understand and control.

- **The algorithm is order-sensitive.** Some AI models, including *recurrent neural networks* (RNNs), work on the basis of 'streaming' data – data is entered continuously, followed by a continuous result. We see such models for recognizing language. This is logical, since the meaning of a word often depends on the surrounding words: there is a *sequence effect.* However, because of this effect, a certain input (such as a word) can be interpreted differently depending on the earlier and later input. The models have a 'memory' which influences the result. [17]

[5] A covariance shift can occur when the nature of the data changes. This means that the data on which a model is initially trained is no longer representative.

Example from the telecom sector: Non-deterministic systems

To embed virtual network components in the physical infrastructure, an AI system can determine the best configuration. There are, however, countless configurations to work out and it is too costly to figure out the best one by *brute force.* Inspired by chess computers, engineers apply the *Monte Carlo Tree Search* to achieve the best configuration. [18]

This method, due to the use of sampling, can never guarantee that the same configuration will be suggested twice, given the same environmental factors. Testing can therefore only provide some level of certainty how the model will behave in different situations. In addition, because the best configuration is not known in advance, it is not possible (for a human being) to verify whether the AI system indeed managed to come up with the best configuration. Of course, comparisons are possible with solutions for different systems or with the results of more traditional optimization algorithms to assess how ’good’ or ‘bad’ the outcome is; for example, the AI outcome could be disregarded if it is worse than from a traditional algorithm.

2.2.3 Who is responsible when things go wrong?

An AI system is not a legal entity but should be seen as a tool (just like a computer) helping a legal entity perform an action. [19] Thus, the use of AI systems does not remove responsibility from the person who decides to outsource their task. The development of such systems usually involves various actors, who each make their own contribution. In such cases, there is uncertainty about who exactly is responsible when tasks are automated.

It is also questionable whether the person or the legal entity using an AI system knows enough about that system to be able to take responsibility for it. This is probably a regulatory and/or supervisory task for the government: although many AI systems are designed to (literally) let people keep a firm grip on the controls, it is very questionable whether a person is also able to recognize when AI goes wrong, or respond in a timely manner. Thereby the effect of habituation also plays a role if an observer sees the AI always making the right choice.

Example from the telecom sector: Who is to blame?

An AI system is trained by a supplier and applied by an operator. The supplier trains the AI system using data from other telecom networks. Who is to blame when a wrong decision is made based on the AI system? In some cases, this can be set out in SLA agreements. AI can also impact key performance indicators (KPIs) that are not specified in these contracts. In such cases, it may be unclear who is responsible.

2.2.4 Autonomous systems can be abused

An AI system’s outcomes depend on the input values. If it is possible to vary this input, it may be possible to change the outcome. With a deterministic algorithm, it is possible to reason how a certain change in input affects the output. As indicated above, this is generally much more difficult or even impossible with AI systems. Consequently, they are vulnerable to so-called *adversarial attacks.* [20] An adversarial attack is based on manipulated data that humans cannot distinguish from legitimate data. The data can be manipulated in such a way that AI makes the wrong decision, without people being able to recognize where the fault lies. [21]

Adversarial attacks, often generated by other AI systems, are therefore usually automated. In theory, there are few AI systems that cannot be fooled by a *threat actor* with enough computing power.

Example from the telecom sector: Adversarial ransomware

An anti-malware system based on AI recognizes files with unsafe content. A malicious party also has this system in place, and has used it to train its own AI (the "adversarial AI"), which can generate files that are unsafe (e.g. contain ransomware), yet not recognized by the anti-malware system. As the anti-malware system does not identify the danger in these files, the ransomware can spread within the network [22]. As a result, a large part of the telecom provider's files is encrypted, and the provider has to face extortion.

Along with manipulating the input data, manipulation is also possible in other stages of the AI lifecycle, as shown in Figure 4. During the training phase, the AI can (ultimately) be made to behave differently by manipulating the training data. Manipulation of the model and/or its structure can lead to the disclosure of information from the training data. The previously discussed *adversarial attacks* happen when the AI is actually in use (the production phase).

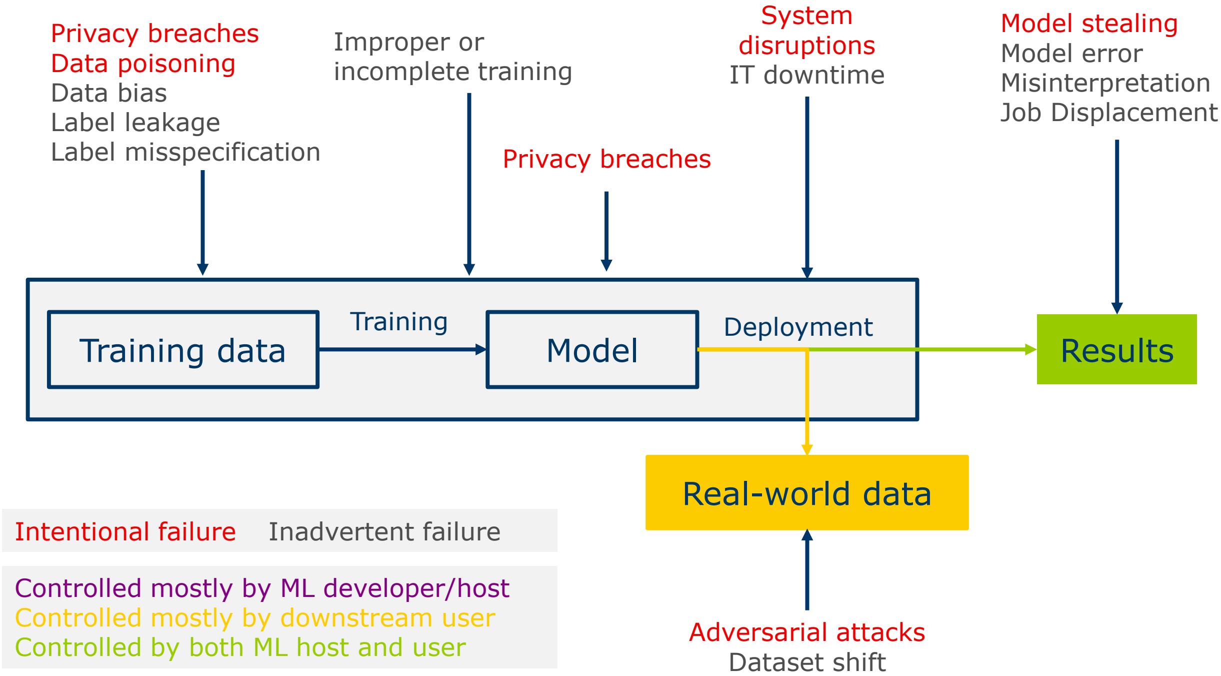


Figure 4 Potential attacks on an AI system during its lifecycle. ML refers to machine learning. [23]

2.3 Current AI applications in telecom

Our research involved exploring the current and future applications of AI in telecom. This was based on agency research (in particular searching for case studies and white papers on product offerings) and discussions with telecom operators and suppliers.

AI is currently mainly used in telecom infrastructures for: 1) configuration, planning and optimisation of how networks function, and 2) maintenance of the network. Below we give an overview of the current applications of AI in telecom infrastructures identified in this research.

2.3.1 Optimisation of telecom infrastructure

The telecom sector has experienced a number of automation phases. Whereas previous connections were still made manually by switching cables, this work was automated by hardware. Now we see these features no longer needing specific hardware but being virtually defined through software.

The heuristics that optimize the network are devised by humans; think of the heuristics that determine how important or threatening a data package is that has to be routed, or how radio resources are allocated to a mobile network. Algorithmic optimization (of a configuration) of functions in telecom networks has also been taking place for quite some time. These algorithms work on the basis of (among other things) traditional mathematical optimization methods. Some of these techniques have been around for hundreds of years. For example, the Newton-Rhapson method (1690) is used to determine the optimum of a mathematical function based on the derivative. The weighted least squares method developed by Gauss (1735) lies at the heart of solving regression problems. Methods such as linear programming (1939) help to optimize a system with preconditions.

A large amount of data is generated in telecom networks, and this data is becoming increasingly available for analysis in one central place. For example, Indian mobile operator Reliance Jio generates 4 to 5 petabytes of data daily from the network's operations. [24] This data is ideally suited for analysis and optimization.

In machine learning, we find a new generation of algorithms that can be used for similar purposes as 'traditional' optimization techniques. The availability of large amounts of data enables the use of machine learning in telecom infrastructures. Machine learning methods are not tied by many of the limitations in traditional techniques. For example, traditional methods can only approximate linear functions, whereas with deep learning, in theory every continuous function can be modelled (the *Universal Approximation Theorem).* [25] In addition, unlike machine learning, traditional methods typically require assumptions about underlying distributions (such as a normal distribution) and input parameters to be independent. Machine learning systems, however, can work without these assumptions.

The following AI applications are currently being used or developed for telecom infrastructures:

- *Power management:* Machine learning is used to achieve power savings in mobile networks. Based on meteorological data, the number of users and their position, antennas actively adjust their radiation pattern, direction and strength to demand. This results in energy saving, for example during the night when data demand is relatively low, and in a more efficient use of the base stations, because a larger surface area can be operated at set-up points where the demand for capacity is not uniform.
- *Radio optimisation:* Currently, machine learning is used to optimize the flow of data to and from a base station in a mobile network. The distance to users, the number of connected users, and certain environmental factors determine the *radio parameters.*[6] [26] They in turn determine the maximum amount of data that can be

[6] Including the form of modulation (in LTE/5G: QPSK, 16-QAM, 64-QAM, 256-QAM, etc.) and the number of bits used for error correction.

transmitted per quantity of spectrum per unit of time (the reason being that with a potentially more efficient modulation, customers with a weaker signal cannot be served again). Interference also plays a role: radio resources can be coordinated between micro and macro cells. To maximize efficiency, algorithms are being used (already) to dynamically determine what part of the spectrum should be used for which user and with which parameters. The parameters of these algorithms can be "tuned" with AI.

- *Quality of Transmission (QoT) estimation*: With optical connections, the signal can be disturbed or interrupted. Machine learning is applied to estimate in advance how well the transmission will work over a connection. Based on things like the cable length, other signals within the cable and the equipment's age, it calculates the best path. The traffic is routed on the basis of this assessment. It is also conceivable that such algorithms are used in wireless networks, for example determining how much error correction or redundancy (e.g. retransmission) is used.

- *Optical network signal amplification*: In optical networks, signal degradation occurs at several points. Currently various AI techniques are applied (such as QoT estimation) to identify points and moments when degradation can take place. Here, the signal is strategically reinforced to minimize noise during transmission.

- *Path computation:* In order to determine the best route between two nodes in a network, several algorithms have been developed that apply certain heuristics. For example, algorithms such as A* or Dijkstra were traditionally used to calculate the *shortest* path. However, there are more factors involved in determining the optimal path, and it is difficult to integrate them in traditional algorithms. Machine learning, on the other hand, takes into account such things as congestion and bottlenecks in the network and thus better estimates the optimal routes.

- *Self-organizing networks*: Based on available information, network components can configure themselves to a limited extent automatically. In this way a mobile base station can find which other base stations are nearby, and thus automatically determine "neighbour relations". Such functionality can be used to quickly organize parts of a network. In practice, however, we see that the functionality is then disabled, and afterwards "fine tuning" is done manually. The functionality generally proves too unstable to allow for dynamic reconfiguration.

2.3.2 Maintaining telecom infrastructure

We see the following applications being used to maintain telecom networks:

Performance monitoring: Monitoring signals within an optical transmission network is essential to detect malfunctions. This is usually done by measuring various parameters, such as *optical signal to noise ratio* (OSNR), *non-linearity factors, chromatic dispersion* (CD) and *polarization mode dispersion* (PMD). By monitoring these variables, problems in the network can be identified in time. Machine learning can better estimate which combination of values increases the risk of interference, and when it is best to intervene.

Predictive maintenance: When equipment fails, this is very costly, both in terms of repair costs as well as lost sales and claims. Many factors affect the wear and tear of equipment and parts, including the weather, the average intensity of use and the type of component. To minimize network outages, models based on machine learning can predict when outages are expected. Preventive maintenance can then be carried out.

2.4 Future applications of AI in telecom

This study, based on our desk research and discussions, also with experts, identified a number of future applications of AI in telecom infrastructures.

2.4.1 Optimisation of telecom infrastructure

In the future we expect to see a number of new forms of optimisation based on machine learning:

- *Smart handovers:* In mobile networks, the signal can decline sharply if the distance to the base station increases or if there are physical barriers between the receiver and the base station *(path loss, penetration loss*). In a classic mobile network, handovers have to solve this problem, but that approach also has its limitations. A future solution to better mitigate this problem is *Multi Tower Beam Forming,* also known as “coordinated multipoint” or CoMP. The signal focuses on a device through a combination of coordinated signals from multiple base stations. Such models can be based on AI and are expected to perform better than models that do not use it.

- *Network orchestration*: Thanks to *Software Defined Networking* (SDN) and *Networking Function Virtualisation* (NFV), it is easier to define a network in one central place and combine data from various sub-systems. This data subsequently serves as input for machine learning models that can optimize the collaboration between the network functions. Several parties are devising a network based on AI. To this end they are gathering large amounts of data from all the network elements (such as measurement data, traffic data and so forth). This data will be processed in a model to create a set of configuration parameters for these same network elements. Based on changes in the network or its use, the AI can quickly reconfigure the network.

- *Optical network nonlinearity mitigation:* In fibre optic networks, noise can occur due to non-linearities. Machine learning can be applied to clean up the signal for further processing. This consequently increases capacity.

2.4.2 AI-based telecom functions

A number of new network functions can be achieved based on AI:

- *Bandwidth slicing / Resource allocation:* In telecom networks, various applications use the same infrastructure and available resources. These applications do, however, have different requirements. Deploying resources efficiently (think of spectrum and “resource blocks” in mobile networks) can meet these requirements optimally. For example, applications that demand low latency work better alongside high capacity applications. The network “recognizes” the low latency requirement and treats the traffic differently than capacity-driven traffic. Machine learning helps to both recognize traffic flows and optimize the distribution of resources.

- *Virtual topologies (VT):* With virtual typologies, not only are the network functions virtualized, but also the overall topology of the network. Thus, we can determine dynamically which transmission paths (light paths) to use, or, where possible, additional capacity should be provided, for example in the form of more connections or an additional local data centre. Machine learning algorithms determine the optimal parameters of the topology. The high speed of this decision-making means humans cannot take on such a continuous configuration task.

- *Anomaly detection / malicious traffic detection:* Here, AI is used to model normal behaviour within telecom infrastructures. Predictions of infrastructure behaviour are then compared to reality: if there is a large discrepancy between the real value and the predicted value, an event is considered deviant. By training anomaly detection on large log files, this technology can be applied in virtually all points in the network to identify deviant events.

- *Dynamic spectrum application:* at present*, licensed* spectrum is mainly statically divided between operators. In systems like the American CBRS,[7] more dynamic, automated mapping is possible. Users first "listen" to which signals they can receive *(sensing),* and/or consult the database (which frequency blocks are reserved?). They can then apply for spectrum use for free parts (automated). Based on algorithms, spectrum will be assigned more dynamically and between operators and other users. This technology was successfully tested in the *DARPA spectrum collaboration challenge.* [27] Such algorithms could be utilized for allocating (licences for the use of) frequencies.
- Nowadays AI is still mainly used for micro-optimizations within specific components or functions. In the future, if current developments continue, AI is expected to play a more central role in telecom networks. This development is in line with the more general trend of (network) virtualization. [28] Controlling networks centrally, and "abstracting" the underlying infrastructure, creates a higher degree of flexibility with regard to the network layout. As AI can be used to implement this central control optimally, it acquires a more *holistic, controlling* role within telecom infrastructures.
- Figure 5 below shows the architecture of a virtualized network that can be controlled by AI, as defined by an ITU working group. [29] The network has several network functions (NF). The functions generate data and make it available for training AI (based on machine learning, abbreviated here as ML). The network functions also support automatic control by AI.

[7] Citizens Broadband Radio Service. See e.g. [wikipedia.org] for a brief description.

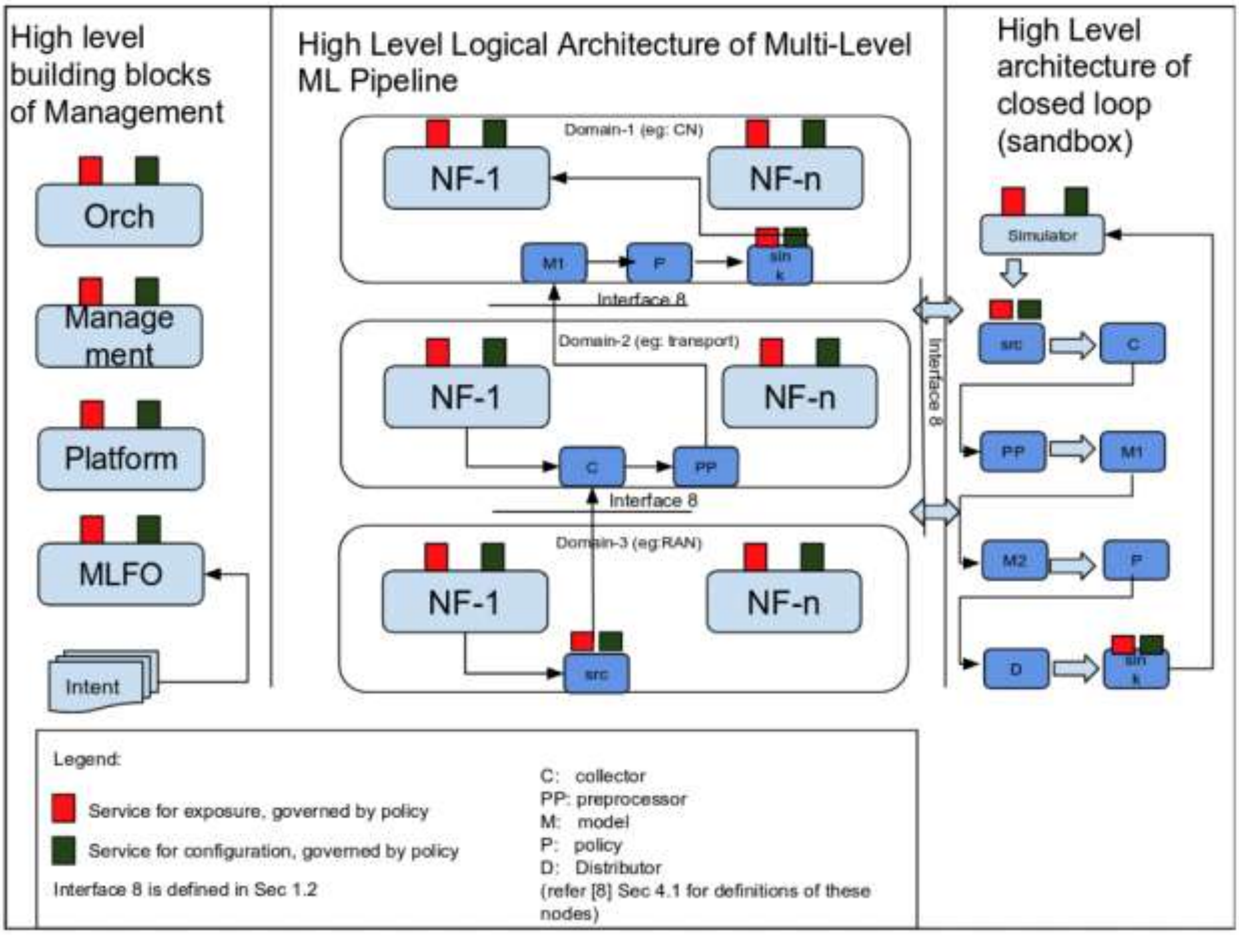

High level building blocks of Management
Orch
Manage ment
Platform
MLFO
Intent
High Level Logical Architecture of Multi-Level ML Pipeline
Domain-1 (eg: CN)
NF-1
NF-n
M1
P
sin k
Interface 8
Domain-2 (eg: transport)
NF-1
NF-n
C
PP
Interface 8
Domain-3 (eg:RAN)
NF-1
NF-n
src
Interface 8
High Level architecture of closed loop (sandbox)
Simulator
src
C
PP
M1
M2
P
D
sin k
Legend:
Service for exposure, governed by policy
Service for configuration, governed by policy
Interface 8 is defined in Sec 1.2
C: collector
PP: preprocessor
M: model
P: policy
D: Distributor
(refer [8] Sec 4.1 for definitions of these nodes)

Figure 5 Architecture for an AI based telecom network taken from an ITU focus group report "Machine learning for future networks including 5G". [29]

3 Risks of using AI in telecom infrastructures

In this section we present a model for analysing the risks of using AI applications in the telecom sector. That is to say, the model provides a framework for establishing a qualitative level of risk based on an application's specific characteristics. On the basis of this assessment, the supervisor can determine whether the risk is acceptable, or whether mitigation measures should be taken.

When assessing the risks of AI in telecom infrastructures, we distinguish the *systemic level* (risk that a telecom infrastructure works as a whole) and the *application level* (risk with an individual AI application within a specific part of the infrastructure).

Figure 6is the risk model developed for this study. At the top is an outline of the systemic level: the people, external factors and applications (including AI) that can cause risk events, which (ultimately) have a negative impact, causing companies and citizens to lose faith in telecom infrastructure and its applications. Our study focuses specifically on AI applications. At the systemic level, it is about embedding these AI applications. At the application level, we examine the relationship between these applications' characteristics and the likelihood and impact of risk events. Later in this section, we will go into more detail about both levels.

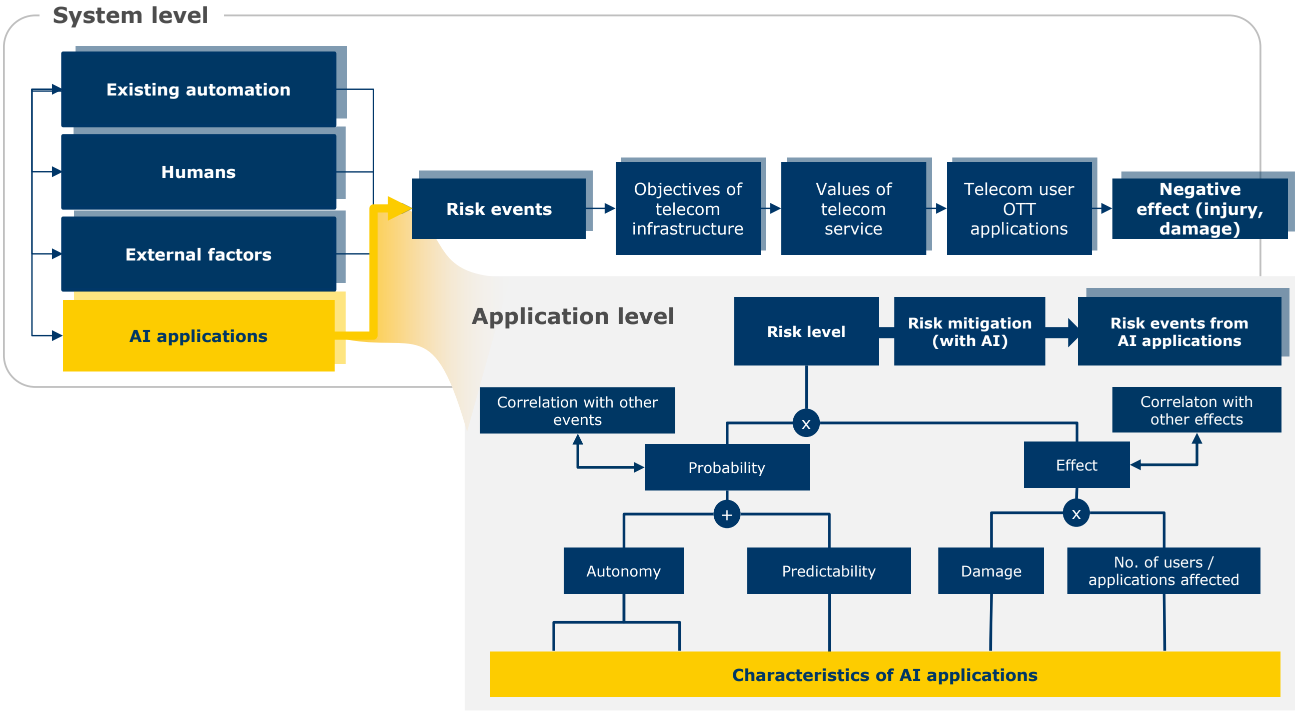

System level
Existing automation
Humans
External factors
AI applications
Risk events
Objectives of telecom infrastructure
Values of telecom service
Telecom user OTT applications
Negative effect (injury, damage)
Application level
Risk level
Risk mitigation (with AI)
Risk events from AI applications
Correlation with other events
Correlaton with other effects
x
Probability
Effect
+
x
Autonomy
Predictability
Damage
No. of users / applications affected
Characteristics of AI applications

Figure 6 A model for the risks of AI in telecom: at the systemic and application levels

3.1 Systemic level

3.1.1 Theoretical framework

Broadly speaking, a risk is a negative event that can occur with a certain probability. Although it is (or seems to be) intuitively clear what the risks are, there is no one single definition for the telecom sector. It is even questionable whether risks can be objectively assessed, or that there are necessarily associated subjective assumptions and choices. [30]

To some extent risks can be assessed in advance. This does not mean that such an assessment is always correct, *complete and objective,* or even possible. It is important to realize that not all risks are *knowable.* To illustrate: an aircraft manufacturer could base the risk of crashing on the failure probability of individual components and the impact of that failure. However, the aircraft manufacturer must also take into account the simultaneous failure of components. These risks are knowable but can of course be "missed". In addition, there are risks that the aircraft manufacturer *cannot* assess: retrospectively, a component could appear to have been sensitive to radiation, yet the aircraft manufacturer had not identified or been able to logically deduce this aspect. A final category, the *known unknowable risks,* involves the risks which people know *can* exist*,* but their precise extent cannot be estimated (Figure 7). In de context of this research, importantly the modelling can only involve *knowable* risks.

	Knowable	**Unknowable**
Unknown	Unknown knowable risks	Unknown unknowable risks
Known	Known knowable risks	Known unknowable risks

Figure 7 The four risk categories

In this study we apply the definition of risk as stated in [30]: A (knowable) risk involves a potential *risk event* (scenario), the *probability* this will occur and the (negative) *effect* of this event. The higher the product probability and effect, the greater the risk. [31] Depending on the method, another weighting or multiplication is used, see figure 5.7 in [30]. Lowering the probability and/or the impact are the logical ways to *mitigate* the risk. Conversely, the risk can be *accepted* if the negative event is highly unlikely, and/or if the negative effects are small or acceptable.

3.1.2 Social risks

Telecom infrastructures play an important social role and are considered vital. [32] More and more services are being delivered digitally, and have thus become dependent on a well-functioning, reliable and always available telecom infrastructure. Users therefore have certain expectations when using telecom networks. Failing to meet these expectations can have negative consequences. We now look at which social objectives a telecom network fulfils and how AI can influence them.

From the perspective of the entire telecom system, we should also consider the *societal effects*. If applications use telecom infrastructure and cannot work properly due to for example outages, society has to pay the costs. In mission-critical situations, there may even be injuries. One example is the 'emergency button' on all C2000 two-way radios used by the police and the fire brigade. If that button does not work, officers in an emergency cannot call their fellow officers in time and may become victims in a dangerous situation. [33]

So what are the relevant target parameters related to any risks in telecom infrastructures? In [30] we see a telecom system model based on services consisting of a network of *nodes* and *links,* both linked to a set of risk events. Our research looks specifically at the effects of telecom service *outages* as a result of these events. In Vriezekolk's dissertation, [30] [30] along with outages, we see other telecom infrastructure objectives that should be included in the research analysis as discussed below.

Availability of networks

Society depends more and more on the availability of telecom networks. As it is becoming such a critical infrastructure, telecom requires the highest possible availability. Applying AI can increase that availability, but also have a negative impact. When an AI system fails, it can sometimes shut down large sections of the network. If errors propagate from system A to system B, this can cause a chain reaction.

Example: Chain reaction

A mobile network applies “power management” procedures based on AI, determining which frequency bands are used in a base station. [34] If there are not many users, the number of bands is reduced to save energy. If it transpires that there have been no users in the surrounding areas for a longer period, the model can disable the base station entirely. Consequently, a system to automatically determine “neighbour relations” (pairs of base stations between which a terminal can move in the network) fails. It seems as if there is less traffic for other base stations. This disables more base stations and propagates the failure throughout the network.

Integrity of information

Integrity is all about the accuracy and reliability of information. AI can have a negative impact on some areas of application. In its informing role, AI can compromise the integrity of information. An AI system may be able to transform noise from another system into incorrect information.

Example: AI-generated information ensures obfuscation

A *predictive maintenance* system uses moisture sensors in the ground to predict when cables will corrode. However, a broken sensor keeps on delivering the same data. The AI model, which is not prepared for the "broken sensor" situation, interprets this data as if the sensor is working. As a result, a corroded cable is spotted too late. One solution would be to run the sensors redundantly or include more (different) information in the model.

Reliability of data

Telecom providers are dealing with large amounts of sensitive information; not only the information that customers exchange over the network, but also (meta)data about customers and this traffic. AI applications in telecom infrastructures can be (partly) trained based on this sensitive data. It is conceivable that a malicious person can trace sensitive data from these AI systems.

Example: trace personal data from AI

An AI system detects fraud on the basis of subscribers' characteristics. The system is continuously trained using customer data and information on fraud cases. It is possible that one characteristic (or specific combination of characteristics) only appears for one subscriber. Everyone with access to the fraud detection system can now trace this particular person as fraudster by entering the data. The system has to be built in such a way that the number of persons who can be traced (the cell size) based on outcomes is always above a certain lower limit.

Potentially unethical choices in telecom infrastructures

When choices have to be made in a telecom network, for example about which traffic has priority or where certain capacity is used, some social *values* may be inadequate or not observed. For example, if emergency services must always be able to use a certain minimum capacity in a mobile network, then this requirement should not be undermined by the application of a particular AI algorithm. Unlike the aforementioned impact, this is an effect at the societal level, where the objective is not a specific application (including provision of the required connectivity) but the socially desirable (ethical) outcome.

In some cases, a decision can be made on the basis of data that should not be used for this purpose or is at least debatable. Several examples are known, such as applications that use the battery level of a smartphone to determine the user's creditworthiness. [35]

Example: Decisions based on undesirable information

AI that has access to all the data available to a telecom operator makes integral decisions based on that data. Because AI itself models the relationships between in and output, it is not known in advance how that information will be used for decision-making. Can AI for example give calls to customer service lower or higher priority based on the fact a customer has not previously reported a complaint to customer services? The same risk applies when telecom data *outside* the telecom sector is used in AI applications.

Example: Pre-emptive or prioritising?

In a disaster scenario, it is crucial that emergency services can always communicate. This can be arranged in different ways in a network. For example, capacity can be permanently reserved for emergency services (a type of "escape lane"), or the capacity of users with lower priority is reduced the moment emergency services need it. [36] In the past, various techniques proved to have individual advantages and disadvantages in different scenarios. A person or system that designs and manages a telecom network must weigh up these pros and cons in a normative framework. An AI system may not be able to make its own appropriate normative decisions.

3.1.3 Risk propogation in the telecom chain

Risks should be considered not only individually, but also in relation to each other: what happens when two events occur at the same time? What happens if one event is the result of another event?

Figure 8 shows schematically how AI applications, people, external factors, and existing automation (whether or not interactive) can lead to risk events. These events negatively impact telecom infrastructure's *objectives*, endanger the telecom services' values, and negative effects may arise due to the malfunctioning of the individual applications. Here we focus on the violation of telecom infrastructures' objectives.

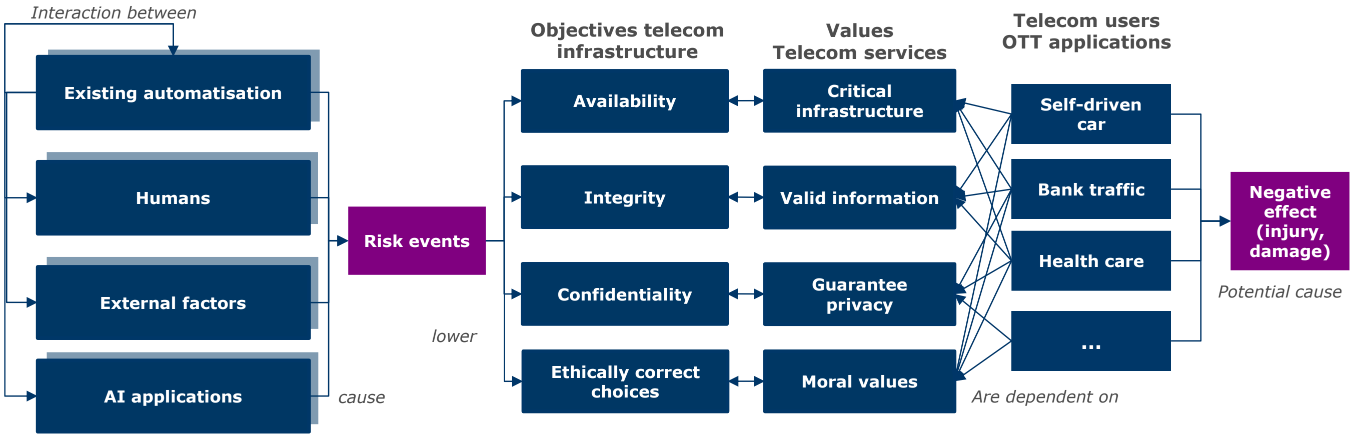


Figure 8 How AI applications cause risk events and ultimately have a negative societal impact

Deploying AI applications in telecom infrastructures can make a *difference* to the overall extent of risk at a systemic level. Any risk that already existed and has not increased or been reduced by applying AI is not considered in this study. Differences can arise (see Figure 8) in the following ways:

- **Applying AI.** We discuss the risks at application level in paragraph 3.2.

- **Interaction between AI application and other systems.** We will also discuss this in paragraph 3.2 correlated to probability and/or effect.

- **Replacing people with AI**. Having AI perform a task incurs risks, and these can be both higher or lower than when a person performs that task. This study does not chart the risks of human action in telecom infrastructures. However, the model presented in paragraph 3.2 can help to determine the risks of the replacement AI application, and thus inform the decision whether to deploy an AI application that replaces humans.

- **Cyber (in)security of AI applications.** Of course, AI applications are also subject to cyber threats and associated security risks, which we discuss in paragraph 3.1.4.

- **Applying AI to mitigate risks.** In contrast, this concerns AI applications specifically geared to mitigating risks. Obviously, the positive effects have to outweigh any potential new risks. We discuss this aspect in paragraph 3.1.5.

3.1.4 Cyber (in)security of AI applications

AI applications are information systems and are therefore subject to all potential cyber threats (breaches of confidentiality, integrity and availability of information). Berghoff et al. [37] provide an analysis of weaknesses broken down into the various phases of an AI application's lifecycle. This highlights the large number of risks that already existed for non-AI-based information systems in telecom infrastructure: information security is needed wherever we use data. The risks found by Berghoff et al. [37] apply to some extent in regular systems and to the decision-makers (and thereby deal with manipulated information). Table 1 is an overview of the new weaknesses found (in our view) when using AI.

Table 1 Potential weaknesses of AI applications regarding information security [37]

Phase	Reliability	Integrity	Availability
Planning	• The (partial) use of existing models that contain malicious elements • Backdoors and bugs in software frameworks for machine learning		
Data collection	• Large amounts of data are required to train a model; this concentration of data is potentially risky.	• "Poisoning attack" whereby training data is manipulated to influence the ultimate result of the AI application (e.g. with hidden "trigger patterns"[8])	• A bias occurs in the training data for a subset of cases, causing the ultimate AI application not to work properly for this subset
Training	• Training is often conducted in a shared (cloud) infrastructure, making it more difficult to guarantee confidentiality.	• Training in a shared (cloud) infrastructure means there is a (greater) possibility of sabotage occurring.	
Testing & evaluation		• Manipulation of the test set can introduce a bias (in the feedback loop to training).	
Operation	• The model can contain 'hidden' information on sources (e.g. if a cell only contains one person), which can expose the model. • Backdoors and bugs in underlying (cloud) infrastructure can endanger reliability.	• An attacker can manipulate weights in a model (thereby introducing 'trigger patterns' and affecting outcomes) without this being detected. • Backdoors and bugs in underlying infrastructure can be inputs for sabotage. • Adversarial attacks.	• Problems found in an AI system are difficult to correct without retraining; the turnaround time and thereby period of unavailability can be unacceptably high.

3.1.5 Mitigating risks based on AI

AI applications in the telecom sector not only cause risks but are also used to mitigate risks. This can be done in a number of ways:

- **Anomaly detection**. Based on small signals or a combination of signals, AI can detect a certain deviation earl(y)ier (e.g. a failing component). Thus, a relevant part can be replaced more quickly, reducing the risk of (later) outages. Another example is using AI-based firewalls that can recognize new forms of dangerous traffic without having previously observed them. This lowers the level of risk because there is a reduced *likelihood* of risk.

  Unlike the use of AI for direct control applications in telecom infrastructure, the risks here are lower. If the system correctly observes certain things ("*true positives"*), the added value is high, whereas not observing things ("*false negatives")* does not lower the risk level compared to the situation without AI. Moreover, incorrectly observing things that are not harmful (false positives) can cause problems, although in many cases these will be less than the added value of the "true positives".

- **Root cause analysis.** In a fault situation or report of failure, AI can be used to determine the cause of failure more quickly. This enables more effective and faster action. The level of risk becomes lower because the negative *effect* is minimalised.

[8] An infrequent combination of input parameters that remains untested in testing/validation, and forces a certain outcome in the AI model.

Simulation. If large sections of a telecom infrastructure are controlled by AI, it is easier to simulate how the system behaves in the event of a calamity. For example, using a copy of the steering model, you can test what happens when large parts of the network fail or for example when misinformation is entered. Such a 'fire drill' can in principle even take place continually. This type of simulation is much more complicated to achieve for an organization where systems work together with people.

When AI is used for risk mitigation, a consideration will have to be made as to whether the deployment of AI caused a net increase or reduction in risk level, and/or the maximum level of risk is not exceeded.

Example: send technicians to the right location

In a telecom network, many network functions are interdependent; the underlying cause of an error may be due to an error occurring in a completely different system. Determining the underlying cause is sometimes difficult, especially (with outage) when it needs to be done quickly. AI can speed up this process by assessing the location of the root cause based on network information. The AI could, of course, be wrong and thus actually delay the recovery process. However, to assess the AI's effectiveness, a large number of simulations can simply be carried out beforehand, to check whether the AI draws the correct conclusion. During a calamity, AI could also make multiple assessments, whereby data from another subsystem is continually removed, and checks are made whether the same conclusion is still valid.

3.2 Application level

3.2.1 Theoretical framework

There are several ways to qualify or quantify risks. A commonly applied method is that of Fine & Kinney [31]. Risk is modelled as the product of *probability, exposure* and *effect,* and these components are scored according to Table 2 below.

Table 2 Method for scoring risk components according to Fine & Kinney [31]

Probability	Exposure	Effect
10 Highly probable	**10** Constantly	**100** Catastrophe, many fatalities, or >$\$10^7$ damage[9]
6 Possible	**6** Daily during works	**40** Disaster, few fatalities, or >$\$10^6$ damage
3 Unusual, but possible	**3** Occasionally (weekly)	**15** Very serious, fatality, or >$\$10^5$ damage
1 Unlikely, but possible in the long term	**2** Every month	**7** Substantial, injury, or >$\$10^4$ damage
0.5 Highly unlikely	**1** A few times a year	**3** Important, disability, or >$\$10^3$ damage
0.2 Almost unimaginable	**0.5** Very rarely	**1** Considerable, First Aid or >$100 damage
0.1 Next to impossible		

The product of the above components indicates the extent of risk and can be converted to a qualitative indication in the table below.[10]

Table 3 Risk scores according to qualitative assessment and measures in Fine & Kinney's method [31]

Score	Risk	Measure
>320	Too high	Consider stopping activities
160-320	High	Apply large measures immediately
70-160	Moderate	Apply simple measures
20-70	Little	Attention required
<20	Slight	Acceptable

Analogous to the Fine & Kinney method, it is possible to determine the risk level of AI applications in the telecom sector by assessing the components' *probability* and *effect*.[11]

3.2.2 Probability

Looking at the aspects of AI applications in the telecom sector that determine the *likelihood* of a negative event occurring, we see a number of categories strongly related to the way the AI application works. Below we explore these aspects in greater detail and show how they relate to the AI characteristics described in paragraph 2.2.2.2

[9] These figures are from an original 1979 study on U.S. Navy weapon systems (merely for illustration purposes and allowances should be made for inflation and context).

[10] See the interactive version at [diasli.de].

[11] The Fine & Kinney method assumes that probability, exposure and effect are knowable and can be determined with certainty. We point out that there may be uncertainty on these axes, and that the most pessimistic value could be used if an assessment of the maximum risk is required.

Autonomy

AI applications often take over tasks normally performed by people, or they support people. This means that these systems have a degree of autonomy. We can distinguish two forms of autonomy: autonomous *learning* and autonomous *action:*

Autonomous learning

Today's AI model has been developed on the basis of large amounts of (historical) data. From this data, an algorithm 'learns' the desired results with certain input. There are several ways to shape this learning or 'training':

- **Offline learning.** A model is trained once or every so often on the basis of a 'static' dataset. Both the model and the data used can be tested and validated before the model goes into production. There are also (non-AI-specific) risks surrounding information security, for example as a result of manipulating the training data, see [37] and para 3.1.4.

- **Online learning**. A model is trained as with offline learning and then retrained periodically based on new data. Continuous testing and validation are also possible. We note that AI outcomes may affect the data used for training, thus creating a kind of 'self-reinforcing effect'.

- **Continuous learning:** A model is continuously updated using incoming data. Think for example of the log data generated in telecom equipment. Unlike with online learning, there are no longer different 'versions' of the model: each inference request has potentially a direct effect on the following AI decision. We see two forms of risks:

    - **(Autonomous) model drift.** Without sufficient supervision, there is a risk that over time, the model will generate incorrect outcomes or present one specific outcome.

    - **Poisoning attack.** An attacker could manipulate the data that the system uses to learn in such a way that the final outcomes also change. An algorithm that should intercept dangerous traffic could slowly 'get used' to such traffic (because an attacker exposes the system gradually to more and more of this traffic), and at some point allow it to slip through entirely. [37]

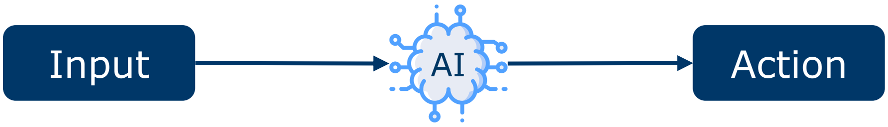
  Berghoff et al. also noted [37] that although an attacker has more opportunities to manipulate where a system is learning continuously, the effects will be temporary.

  The likelihood of negative events occurring as a result of applying AI is greater if there is insufficient testing during an AI model's training phase or validation of whether the applied data and method are adequate or not. This risk is greater when *online learning* is involved and greatest with *continuous learning.*

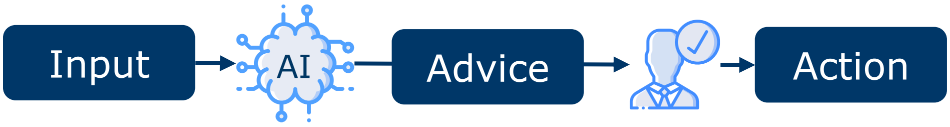
Act autonomously

An AI application can be applied in various ways:

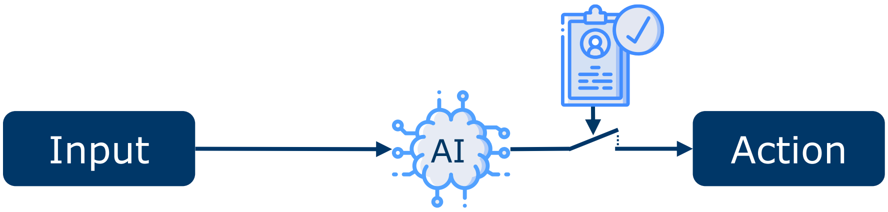
- In a **closed loop** scenario, the AI system performs actions directly. The only action people can perform is switching off the system. One example is speech recognition software.



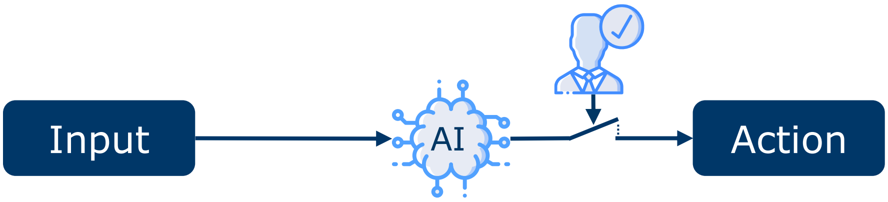
- In an open loop scenario, the AI system's role is to provide support. AI presents a person an outcome, and on the basis of this outcome, the person can act. In this scenario, it is possible for people to deviate from the advice and/or check this advice based on other information. Examples are expert systems that help doctors form a diagnosis.



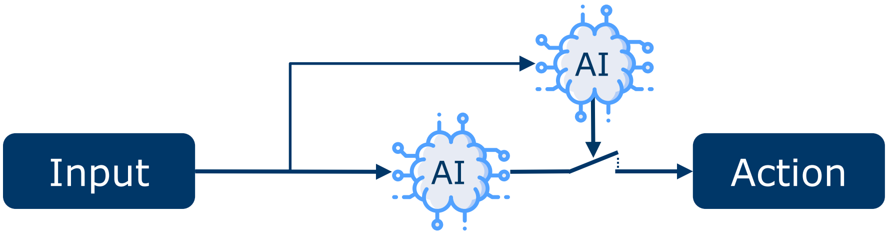
- In a rule-constrained closed loop scenario, an AI system can perform direct actions, but is restricted by certain "hard" rules. Breaking the rules results directly in the system being disabled or failing to perform the action. An example is autonomous vehicles, which are often equipped with various 'fail safe' rules that ensure a car makes an emergency stop in unsafe situations.



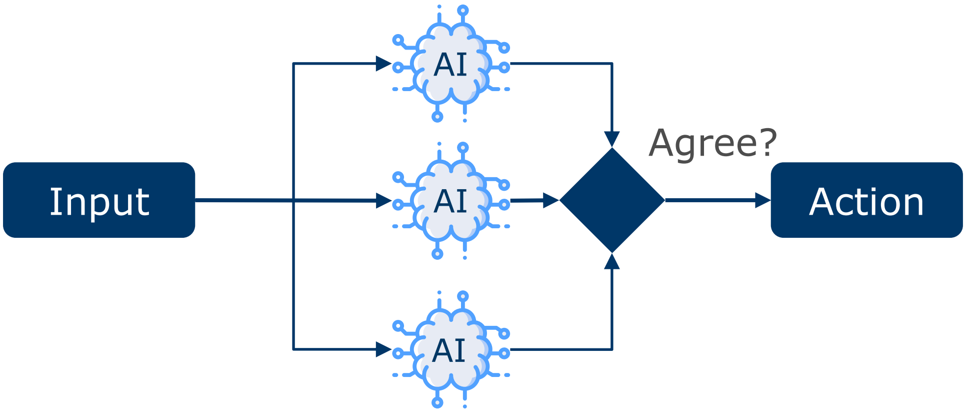
- In a **human-in-the-loop** scenario, AI can perform actions directly, but a human can stop or adjust these actions if necessary. An example is autonomous vehicles where people have to keep their hands on the steering wheel.



- In an **AI-in-the-loop** scenario, one or more additional AI systems monitor an AI system that performs actions. The controlling AI model can view the original inputs and the AI's decision, and assess whether this decision is correct.



Another form of implementation is by having multiple AI systems make the same decision, and only execute it if the decisions are the same. This principle is applied to navigation systems in aircraft. Three computers with different implementations of the same algorithm calculate navigation parameters, and the outcomes are only used to control the aircraft if the three results are exactly the same.



Note that in scenarios involving people ("human-in-the-loop" and "open loop") there is a risk of habituation. Over time, people's trust in AI can grow and/or attention can decline (*"vigilance decrement"* [38]), making it less likely these deviations are detected, and there is in fact a *closed loop,* and therefore a greater risk.

In fully autonomous scenarios, the impact of traditional information security risks is growing: an attacker who can change the weights of an autonomous AI model can remain undetected for longer due to the complexity of the models.

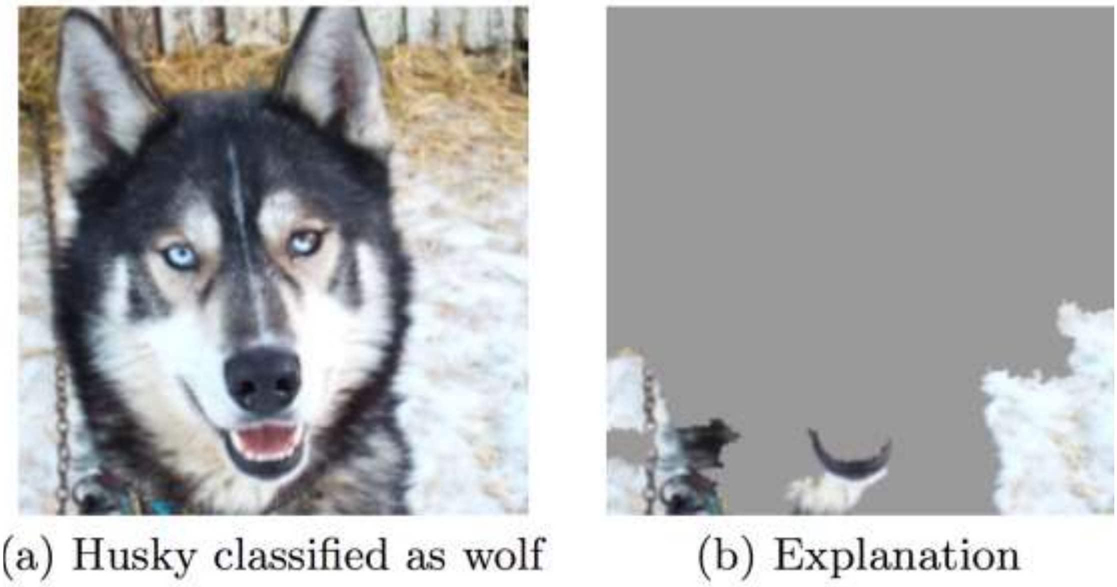
The likelihood of negative events occurring as a result of applying AI is greater if an AI system can act directly. Although the risks can be mitigated by human control, it is highly questionable whether a human can always oversee the consequences of a decision and intervene quickly enough, and whether, over time, there will be too much trust in AI systems. In some situations, AI can perform better than a human, but even then, the likelihood of negative events increases if there is no supervision.

Unpredictability

The degree of predictability has a major impact on the extent to which we can assess the probability of negative effects. As mentioned earlier, AI applications based on *Deep Learning* are much more unpredictable than rule-based algorithms due to the high complexity of such AI models. It depends on the specific form of deep learning and its implementation in the application; the elements we discuss below have a major impact on predictability.

Transparency

Some forms of AI, especially those based on *deep learning,* consist of a large number of layers and coefficients. It is not easy to deduce from these how the model behaves and on what basis decisions are made. Figure 9 presents a striking example. A model was trained to distinguish various animal species. An evaluation of the model showed that although the outcomes were correct, the model's decision to classify an animal as a wolf was apparently based primarily on the presence of snow in the photo. Consequently, new pictures of other animals with a lot of snow in the background also produced the classification 'wolf'.



(a) Husky classified as wolf

(b) Explanation

Figure 9 AI's use of information to classify animals [39]

The lack of transparency in AI models makes it easier for a malicious person (within or outside the organization) to perform manipulations and causes these to remain undetected for longer. In this way an attacker with access to a model can change weights without this being noticed directly, but which create a trigger *pattern* in the system.

A non-transparent system increases the risk of *adversarial attacks.* [40] If the AI is non-transparent, characteristics that are not robust (such as snow in the husky-wolf example) may be used for classification. If it is not known which precise characteristics are being used, a malicious person can manipulate the non-robust characteristics without the system developer being aware of this.

There are also conceivable scenarios where transparent AI is *not* desirable. If a security system has precisely known rules, an attacker can search for holes and the *adversarial attack* is easier to carry out. In a continuously learning non-transparent system, the systematic search for a leak is more difficult, but of course the system is not by definition safer.

Non-linearity

AI models, depending on their implementation, can exhibit strong non-linear behaviour. On the one hand, this ensures that these models can form very powerful representations. On the other hand, such behaviour makes it more difficult or impossible to assess the likelihood of negative effects. Figure 10 is an example of a model where an outcome is assessed on the basis of two parameters (x and y, shown as red and green; this could be a classifier that categorizes network traffic based on two properties like 'good' or 'bad'). As the image demonstrates, the transitions at certain points are sharper than at others. In the centre, the outcome is most sensitive to input changes: a small adjustment can cause the outcome to go one way or the other.

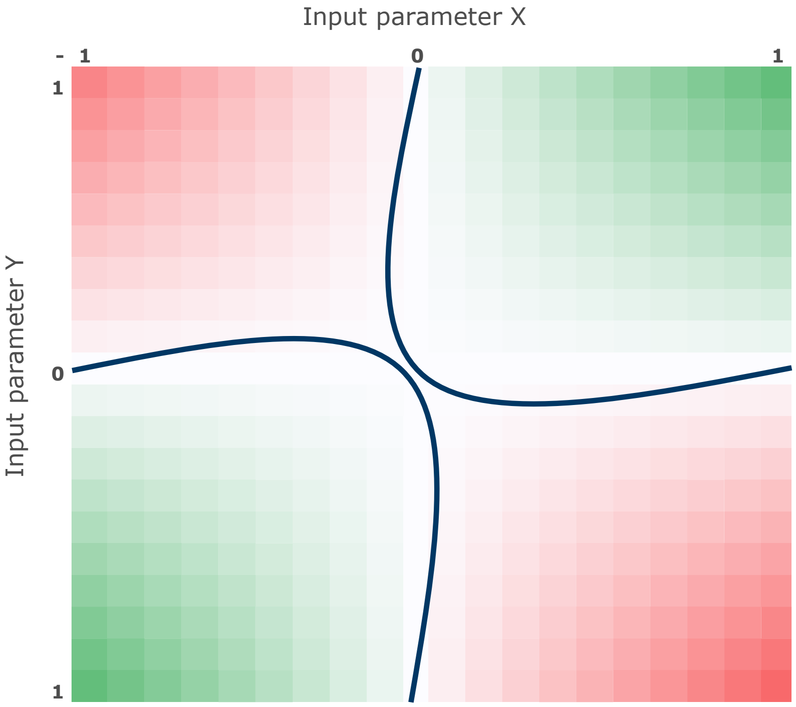


Figure 10 Example of a non-linear model with two parameters and 'decision boundaries' for classification in two categories (source: Dialogic).

Researchers recently discovered an example of nonlinear behaviour in practice, namely in an algorithm for self-driving Teslas that recognizes speed signs. By extending the middle leg of the number "3" with a few centimetres of tape on a speed sign marked "35", the car suddenly identified the number on the board as "80". [41] Figure 11shows this schematically. An AI model's assessment ("probability of it being an 80 sign") increases non-linearity as a result of a very specific characteristic.

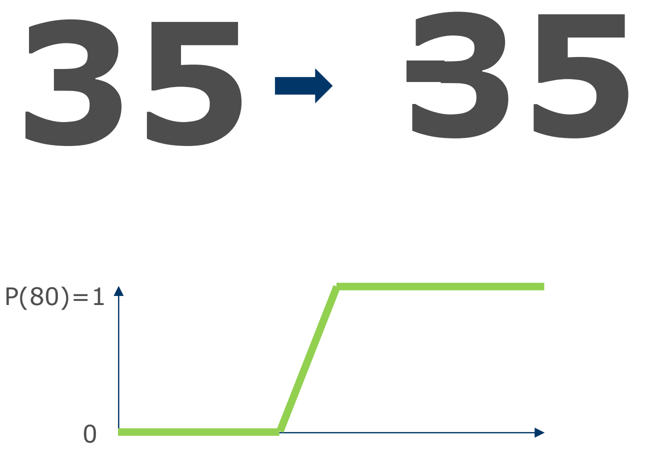


Figure 11 Non-linear activation functions in AI models lead to non-linear behaviour in the model (source: [41], visualisation: Dialogic)

Non-linearity has a greater impact on the likelihood of negative effects if the data the model uses can be manipulated by third parties (as in the example with the speed sign), and if the data is not properly validated and certain input values fall beyond the limits to which the model was trained.

The predictability of AI models has a direct impact on the amount of negative effects occurring as well as the extent to which these can be determined with certainty. The greater the lack of transparency, the greater the risk (the model's operation is difficult to control) and there is the possibility of non-linear behaviour. The risk is greatest if the data can be manipulated or is insufficiently controlled.

There are various methods that increase the predictability of AI systems. One method is to create a simulation that can test the AI system before it is applied. For a classification model with two parameters (x,y) as input that can assume values between -1 and 1, it is easy to explore the entire input-output space. By placing each input parameter on an axis and trying out the potential values, we can determine the *decision boundary (*the moment when the model chooses between A and B) of a classification model.

Because of the large parameter space in which AI systems make decisions, it is difficult to explore the entire input space. [37] This soon makes it very complicated to interpret the *decision boundaries* in a high dimensionality with sometimes more than 1000 input variables.

New situations

An AI model is often trained with large amounts of data collected in the past. Thus, an AI model learns which outputs are suitable for which combinations of inputs. Because this learning is done on the basis of historical data, the AI model assumes that *new*, previously unseen combinations of inputs, can be predicted based on earlier combinations. In some situations, this assumption may be incorrect. For example, it has been shown that historical stock prices can be perfectly predicted based on AI models, but these models are anything but capable of predicting future stock prices correctly. The adage *past results do not guarantee future performance* therefore also applies to using AI.

AI cannot cope well with new situations because it lacks an 'understanding' of the underlying relationships. AI only looks at input and output, and the underlying relationships are nothing more than a 'black box'. The likelihood of risk events increases in scenarios where new situations can arise.

Because of the above, it is important that the limits placed on inputs to an AI model are known and observed. For example, a model could be trained and tested within a certain range of a particular input variable. Technically speaking, such a model will be able to generate outcomes beyond this range (illustrated in Figure 12). However, these outcomes might bear no relation to reality, because the model has never been trained in them: the model 'extrapolates' reality but in a purely mathematical way, without it being established as valid.

AI models have limitations when it comes to input data. If the data entered lies outside the validated range, the outcomes may also be invalid. It is important that these boundaries are known and enforced in the use of AI. Otherwise, the likelihood of risky events (due to incorrect outcomes) increases.

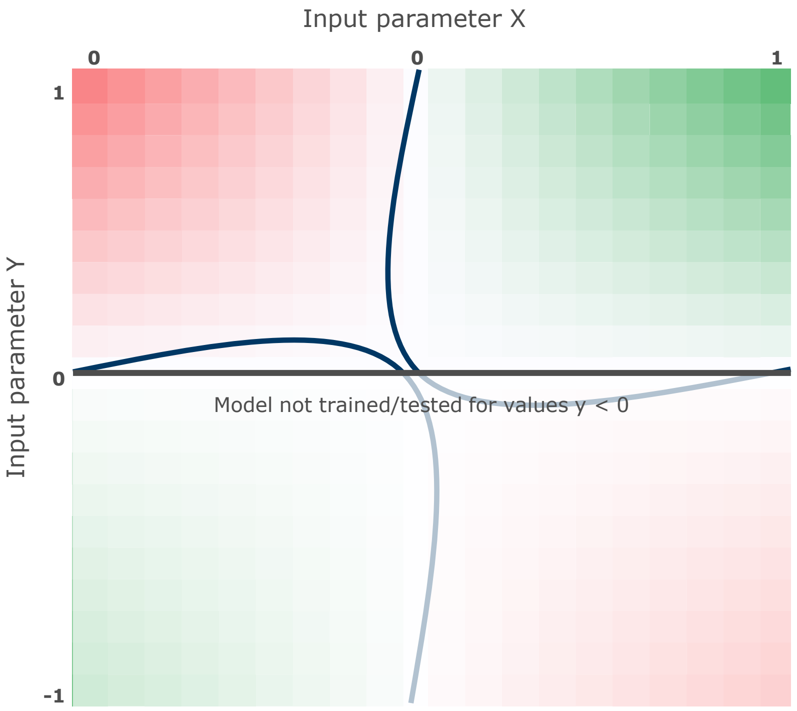
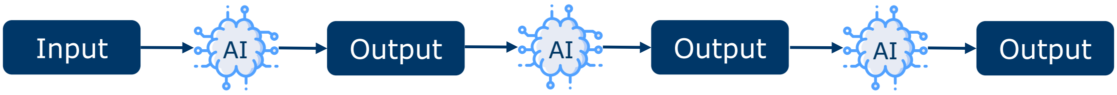


Figure 12 Example of a model trained in a certain input range, that will generate output (the colours in the chart) for values beyond this range

Correlation

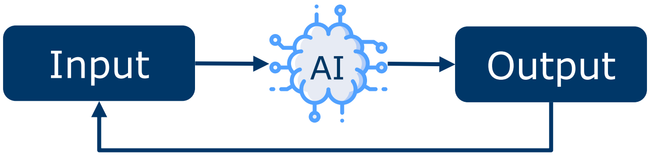
It is conceivable that several events have to take place first before a negative effect occurs. For example, many systems in aircraft are duplicated. In principle, negative effects only occur when both systems fail. However, the events that (together) have a detrimental effect may be correlated. In the aircraft example, it is feasible that both systems have the same defect or suffer from common cause failure. If events are related, we call this a correlation between the events (regardless of the cause). In the context of AI applications in the telecom sector, we see that a correlation of events can impact the likelihood of negative effects in two ways:

- **The output of one AI system is used as input by another.** A fault in an earlier system can thus lead to a fault in a later system, via all the mechanisms described in this paragraph (e.g. invalid input data).



An AI system's output is also used as input for the same system, or the systems are otherwise linked.[12] For the same reasons as above, this can lead to an escalation of incorrect outcomes.

[12] See [26] for an example outlining a situation where several AI "agents" operate and learn autonomously, but share observations.



In a telecom network, a correlation scenario might be: a base station in a mobile network incorrectly stipulates that the radio signal must be amplified in a certain direction. A second basic signal measures an adjacent signal and adjusts its own configuration accordingly, which results in a similar error, and this is detected by the next base station. The error then infiltrates like an oil slick through the network.

Correlation of events can increase the likelihood of negative effects. The highly linked systems in telecom networks mean that when AI applications use each other's or their own output as input, the probability of correlation is highest.

3.2.3 Effect

Looking at the effects of risk events arising from AI applications, we observe two determining factors: *damage* (the severity of the effect) and *scope* (the scope of the effect). In addition, we see that effects can be correlated; in other words, they can strengthen each other if they occur simultaneously.

Damage

The *potential* damage that AI can cause in a risk event is related to the algorithm's action framework (i.e. humans, who would make the wrong decision based on AI advice) – the more important the decisions, the greater the risk. A larger number of different action options means that evaluating them could also be complicated for an algorithm supervisor.

The potential damage is greater if a human cannot (timely) intervene or if the algorithm's decision scope is not otherwise restricted. In this context, all the previous considerations regarding working autonomously (p. 33 apply.

The more influential the (indirect) decisions of an AI application, the greater the negative effects in a risk event. Where an AI application acts autonomously, the negative effects are in some cases greater.

Scope

In addition to an AI application's action framework, the *scope* of AI actions is significant. In a telecom network, the scope can convey the number of (potentially) affected users or geographical areas. A telecom network has different 'layers' (such as access, transmission and core levels) where the scope is constantly expanding. Regarding the scope of an AI application in a telecom network, we see the following gradations:

- **Completely isolated.** The algorithm makes choices that have an impact in a tightly defined environment. The AI outcomes have no impact whatsoever on other systems within the telecom infrastructure. An example is an algorithm that optimizes *beamforming* in a mast or improves noise reduction in a bundle of VDSL lines. A wrong outcome only impacts the connections in question. The applications are generally highly *decentralised*. The scope is limited to (1) a single or small group of users, (2) a geographically strictly defined area, and/or (3) only the access part of the network.

- **Partially isolated.** An algorithm's action framework is well defined, but there are ways an incorrect decision can impact other systems. This occurs, for example, if an algorithm output has a measurable effect on another system. Yet another conceivable route is that the failure of an AI system leads to surpassing security limits (e.g. a residual-current circuit breaker or fuse) that cause other systems to fail. The scope is limited, but is available to a larger group of users, larger geographic area, and/or more than just the access part of the network.

- **Not isolated**. These are systems designed to control other systems. An error in the steering system has a direct impact on the functioning of the controlled systems. The scope is potentially the entire network, all users, and the entire geographic coverage area.

A technique that deserves special attention is *edge computing,* whereby intelligence (be it application-specific, and possibly based on AI) is applied to the periphery of the network. Although these applications' sphere of influence is local, there is a risk of influencing other applications that use the same infrastructure.

In AI applications decentralised at the local level (and risks not correlated with each other), there are generally smaller negative effects on risk events than with central AI applications designed to control other systems.

Correlation

Earlier we discussed that the correlated probabilities of risk events can lead to new (or a higher than expected probability for existing) risks. Correlation can also increase risks regarding effect. An analogy is securing a building from burglary: if the alarm is not switched on, there is no immediate increased risk of damage from break-in; after all, the door is locked. If only the door is not locked, there is also no immediate increased risk (after all, the alarm still works). However, if both the alarm is not turned on and the door is not locked, the risk is much greater than the sum of both risks: a burglar can now enter without any problems, and there is damage.

Domino effect

If systems are connected and decisions in one system affect another system, there is a risk of a "domino effect": a system error causes an error in the next system. Such a mechanism was at the root of the "flash crash" on Wall Street in 2010. One algorithm detected erroneous input as an anomaly and decided to sell securities. Other algorithms saw this act as deviant and acted in the same way, resulting in share prices ultimately collapsing and trading having to be shut down. [42] The rapid actions of the algorithms and the fact that it was not known that the algorithms contained "anomaly detection" mechanisms, meant that share prices slumped very quickly. It is of course equally possible that human stockbrokers would be susceptible to the same thing: they can panic, with ultimately the same effect.

Redundancy and variation

Redundancy is one way of mitigating risks. By duplicating elements, the failure of one element can be absorbed by the other. In addition, the output from the two elements can be compared and when a difference arises, this is noticed (works best if the elements are completely different implementations of the same function). Redundancy is often installed in telecom networks; for example, networks are set up in rings, so that disconnection does not lead to a complete loss of connectivity between locations.

One concept associated with redundancy is *variation.* Introducing redundancy to mitigate risks is only effective if the failure of the redundant elements is not mutually correlated. Two

AI systems that are redundant from each other will – if they are otherwise exactly the same – make precisely the same error given the same input; in that case, in practice there is no reduced risk at all from redundancy. This can be solved by introducing variation. The two redundant AI applications are two different deployments (completely separate implementations, or perhaps an older version).[13] There is therefore less chance that both systems encounter an error at the same time.

The negative effects of a risk event are smaller with an AI application acting in a redundant part of a telecom infrastructure, as long as there is no correlation between preventing risk events in multiple redundant setups.

3.3 Determining risk

As explained earlier, the actual risks of applying AI in telecom infrastructures can only be determined by analysing them at the systemic level. Not only AI applications in isolation, but also the interaction between AI applications and other systems have an impact on the occurrence of risk events. The resulting negative effects are strongly linked to the ultimate use of the telecom infrastructure. It is beyond the scope of this research, and in our view not possible, to create a fully comprehensive risk model, without looking at *specific* applications and situations.

That said, it does make sense to look at the characteristics of AI applications and their direct impact. We call this the *application level*. Figure 13 presents the model for assessing the additional risks incurred with *individual applications* of AI in telecom infrastructures.

As discussed, the level of risk depends heavily on the characteristics of the AI application, which (indirectly) determine both the probability and the impact of risk events. Policies could be implemented for the relevant characteristics of AI applications.

Risk mitigation by the operator can reduce the risk level of an application to an acceptable level (for the operator and society). In this area too, policy can be implemented. For a supervisor, this poses a key policy question: what level of risk is acceptable to society, and what mitigation measures should operators take?

Scoring risk aspects at application level

While the exact weighting of aspects in the risk model may be subject to discussion, based on our research (in particular the literature and discussions with experts), we provide the first steps for a scoring model as shown in Figure 14.

In Figure 14, each risk-relevant characteristic of an AI application is given a score between 1 (lowest) and 10 (highest). The criteria that lead to a specific report score are shown in the grey blocks, as explained in detail above.

[13] It is interesting to see how this is done in the Space Shuttle [space.stackexchange.com].

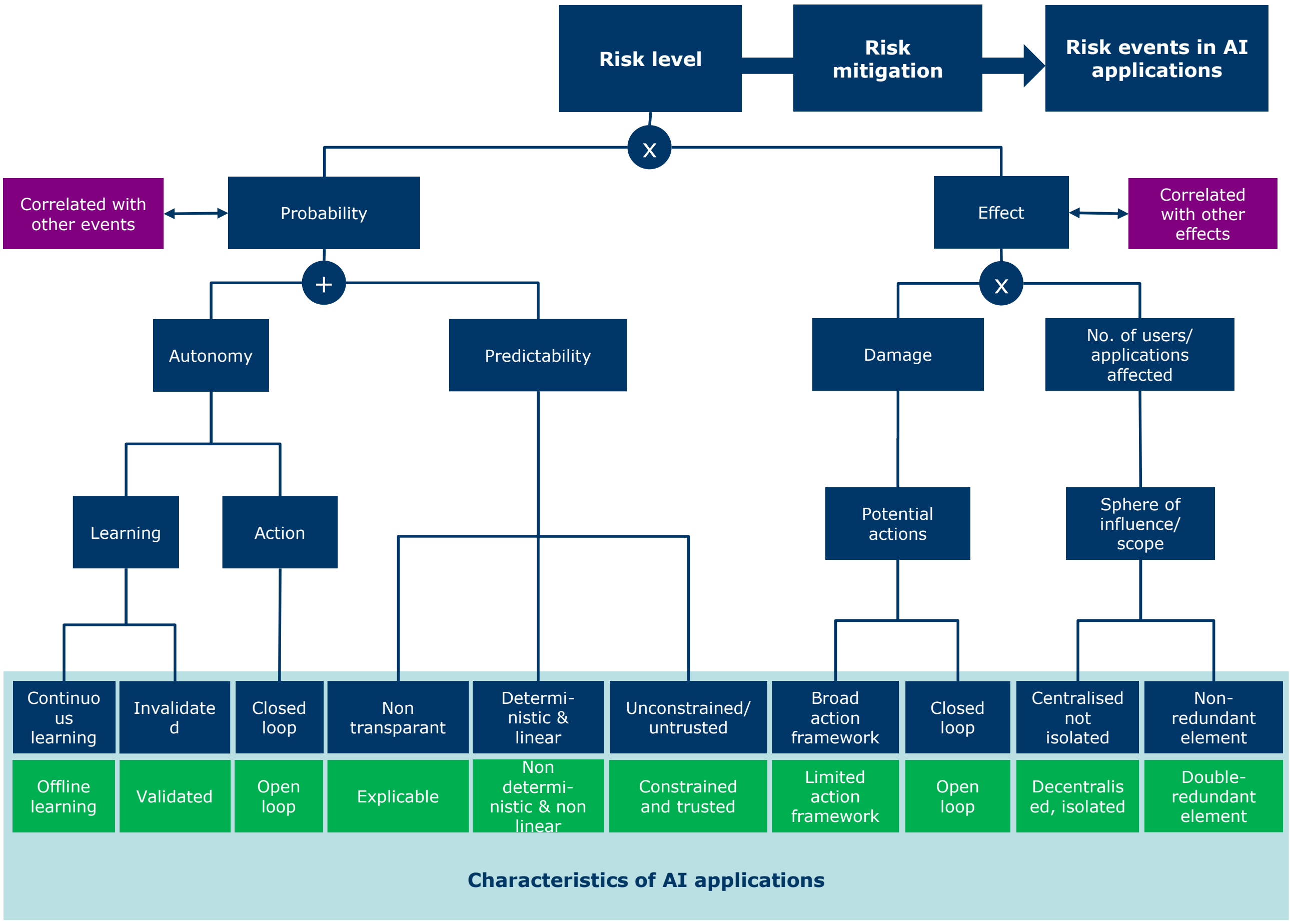


Figure 13 Model of additional risks when applying AI in telecom infrastructures, with policy starting points

Autonomy			Predictability			Damage		Scope		
Learning	Validation	Action	Transparency	Deterministic & linear	I/O	Action framework	Use	Scope	Redundancy & variation	Score in risk model
Continu lerend	Unvalidated	Closed loop	Intransparant, 'black box'	Non deterministic & non linear	Unconstrained/ untrusted	Broad action framework	Closed loop	Central	Non-redundant element	10
●	Model not seen nor tested	AI in closed loop	Model not seen nor certified	Stochastic AI-algorithms	Use unlimited data set	Network set-up	AI in closed loop	Network orchestrator	●	
●		●	●	●	●	●	●	●	●	
●	●	Constrained closed loop	High number of para-meters	Sequence sensitive (RNNs)	Use 3rd party data	●	Constrained closed loop	Edge	Redundant element	
Online lerend	●	●		●	●	Control network element	●	●	●	
●	●	●	●	Non-linear elements	●	Control traffic	●	Base station POP, MDF	●	
●	A few scenarios tested	Human in closed loop	●	●	Use own network data	●	Human in closed loop	●	●	
●		●	Training data unknown	●	●	●	●	CPE	●	
Een-malig getraind	All input combinations tested	AI in open loop	Certified	Linear regression	Only generated data	Optimalisation of parameter	AI in open loop	Handset, terminal	Redundant varied element	
Offline lerend	Validated	Open loop	Explicable 'white box'	Deterministic & lineair	Constrained & trusted	Limited action framework	Open loop	Decentralised, isolated	Double redundant, varied element	1

Characteristics of AI applications

Figure 14 Assessment of AI applications in telecom infrastructures based on the risk model

4 Role of the Dutch Radiocommunications Agency

In this section we discuss the role that the Dutch Radiocommunications Agency as supervisory body can fulfil when dealing with the (additional) risks arising from the use of AI in telecom infrastructures.

4.1 Action framework

In paragraph 3.1.3 we discussed the potential negative social effects (including injury, damage and loss of faith in the system) of applying AI at the systemic level. The supervisory body can intervene at two different points in this process (see Figure 15):

1. **Within telecom infrastructures.** The supervisory body can take measures in order to ensure that no irresponsible AI systems are applied in telecom infrastructures with negative characteristics that (could) lead to additional risks.

2. **Outside the telecom infrastructures/in telecom users' OTT applications.** The supervisory body can try to develop the relationship between applications and demands for telecom infrastructures, in order to limit "irresponsible" use of telecom infrastructures. If the application side has more knowledge and awareness of the risks in underlying telecom infrastructures, a better assessment could be made of the risks at application level.

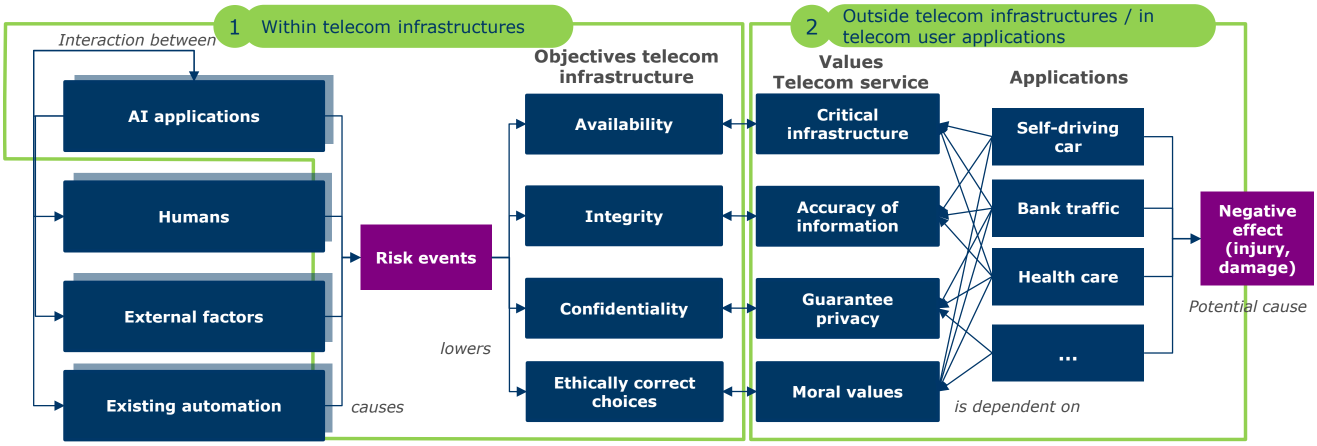


Figure 15 AI characteristics that propagate negative social effects

Demands can be placed on applications and the actual peripheral equipment. Existing European regulations provide a useful framework. The relevant EU directives state that telecommunications peripherals must meet 'essential requirements'. For example, Directive 2014/53/EU, known as the Radio Equipment Directive (RED) states: "*[Radio equipment belonging to certain categories or classes must be constructed in such a way as to meet the following essential requirements: radio equipment does not harm the network or its operation or abuse the network resources causing an unacceptable deterioration of the service.]*" [43] A supervisor could make efforts to 'activate' this additional obligation (whereby the European Commission is authorised to adopt delegated acts that enable such 'activation'). [44]

This study focuses on a telecom supervisory body, and therefore primarily the first link in the chain. That is not to say the supervisory body could not also operate in the application domain. In a sense, the Dutch Radiocommunications Agency is already part of this domain through its *Telekwetsbaarheid* (tele-vulnerability) programme, which raises awareness about outages among users of telecom infrastructures. [45]

4.2 Assessing risks

The way negative effects are considered relates to the level of ambition regarding telecom infrastructure and is therefore a social consideration. If the starting point is that (certain) telecom infrastructures can be used for mission-critical applications (with added social value), then the requirements regarding risks are higher than with a lower ambition (e.g. only business-critical use). We can therefore define risk effects in relation to telecom networks' desired service levels (indeed the level of all the infrastructures combined, because for critical communication, a combination of various infrastructures can spread the risks).[14]

A supervisor aiming to assess AI risks in the telecom sector will first have to determine a level of ambition with regard to infrastructures: if it is socially desirable that mission-critical services can operate on the basis of the available telecom infrastructures, the risks take precedence over when merely 'best effort' services are required.

The previously developed risk model can then form the basis for a supervisory body to determine at application level which AI applications pose risks. It also explains how AI could actually contribute to reducing risks. A supervisory body can consequently regulate specific applications. Another method is to examine the AI applications' characteristics in order to determine which combinations of characteristics lead to a (in the eyes of the supervisory body) too high (additional) risk, and compile regulations on the basis of these combinations.

4.2.1 Use in identified applications

The risk model can be used for the AI applications identified in this study to get an idea of which ones could cause additional risks. Table 4 provides an overview.

At first glance, the assessment in Table 4 might, wrongly, seem to suggest that it would be better to exclude certain applications from telecom infrastructures. We stress, however, that this table only considers the *application level*. As indicated, there may be additional risks as a result of embedding the application within the telecom infrastructure (due to correlation of risks). An AI application can also be used to mitigate risks. Finally, no weighting is given to the scores nor linkage with the negative consequences: the model is not normative with regard to the acceptable level of risk nor the scores in Table 4.

The model does however identify where AI applications with which characteristics can lead to new risks. We can also surmise from the table that more attention should be paid to validation: risks in AI applications that generally have a limited risk could be further reduced.

[14] Not surprisingly, the ISO31000:2009 standard provides a broader definition of the concept 'risk' as ["the effect of uncertainty on the possibility of achieving a company's objectives"] thus recognising that this is not just about the risks that can lead to injuries or damages, but also the risk of not achieving objectives that may have serious effects further down the chain.

Table 4 Scored risk aspects of current and future AI applications in telecom infrastructures[15]

Application	Opportunities: AI-mitigation[16]	Autonomy			Predictability			Damage		Scope	
		Training	Validation	Action	Transparency	Deterministic & linear	I/O	Action framework	Use	Range	Redundancy & variation
Power management		1-6	1-10	7	5	5	4	2-5	7	5	5-7
Radio optimisation		1-6	1-10	7	5	5	4	2	7	5	5-7
Optical network signal amplification		1	1	7	5	5	4	2	7	5	5-7
Path computation		6-10	1-10	7	5-10	5-9	4	10	7	10	5-7
Self-organizing networks		6-10	1-10	10	5-10	5-9	4-9	10	10	10	5-7
Performance monitoring	✓	1-10	1-10	1-4	5-10	2-5	4-9	2	1-4	10	5-7
Predictive maintenance	✓	6-10	10	1-4	5	2-7	4-9	2	1-4	10	5-7
Smart handovers		1-6	1-10	7	5	2-5	4	2-5	7	7	5-7
SDN/NFV		6-10	1-10	10	5-10	2-9	4-9	10	10	10	5-7
Optical network non-linearity mitigation	✓	1-6	1	7	5	2-5	4	5	7	4	5-7
Bandwidth slicing / Resource allocation		6-10	1-10	7	5-10	2-9	4-7	5	7	5	5-7
Virtual topologies		6-10	1-10	10	5-10	2-9	4-9	10	10	7	5-7
Anomaly detection / malicious traffic detection	✓	10	10	10	10	10	4-10	2-5	10	5-10	5-7
Dynamic spectrum application		6-10	1-10	7	5-10	2-9	4-9	7-9	7	7	5-7

[15] The cells are coloured as follows: green: 1 to 4, orange: 5 to 7, pink: 8 to 9, red: 10. In each range, the last applicable colour in the series is used.

[16] The column "Opportunities: AI Mitigation" indicates whether AI is specifically used in this application to mitigate other non-AI risks.

4.3 Tools

The Dutch Radiocommunications Agency could use various tools to reduce the additional risks of applying AI in telecom infrastructures:

- **Information and awareness raising.** Thanks to campaigns in and beyond the telecom sector, operators and users have become aware of the additional risks posed by AI-based systems. Not only can both groups consider mitigating risks, a dialogue can also be initiated to align the levels of services and objectives with the needs of end users.
- **Requirement for transparency.** Operators can be required to provide insight into the use of AI in the network and the way risks are mitigated. The supervisory body could propose a model or format for this. This "risk label" makes it clear to end users whether the network is suitable for the relevant application.

Is certification a guarantee for transparency?

Demands for products are made in many sectors outside AI. In these cases, the manufacturer endorses through a statement (a "supplier declaration of conformity") that the product meets the requirements. IBM researchers argue that a similar statement or certification could be used to improve the reliability of AI*.* ***[14]***

The proposed method would assess the aspects of fairness (and balance in the algorithm), explainability (of the results), robustness (including non-linear effects) and lineage (origin of training data). The manufacturer would have to complete a comprehensive questionnaire on the product.

- **Facilitate risk analysis and mitigation.** The supervisory body could facilitate knowledge sharing about the risks of AI applications in telecom infrastructures. This could be in the form of organizing a dialogue between the various operators. The supervisory body could also speak to individual parties to ascertain ongoing developments and how risks are addressed.
- **Develop criteria.** The supervisory body could develop or designate criteria regarding the use of AI in telecom infrastructures. These could be generic criteria (how to deal with training data, control of autonomous systems, et cetera) or specific criteria associated with certain applications. Various initiatives have been instigated internationally to establish criteria (see [46]).
- **Establish process requirements for operators.** The supervisory body could require operators to establish processes for mitigating additional AI risks, for example by stipulating certain checks or a degree of validation/transparency.
- **The European Commission is proposing and supporting proposals for the 'activation' of additional obligations for telecom peripherals**. As indicated above, the European Commission could adopt a directive to establish additional requirements that help prevent the harmful effects of using AI in peripherals.

The above tools can be used at the systemic level. This is about providing information or establishing transparency requirements at a functional level, related to telecom infrastructure values.

These tools can also examine specific risk characteristics of AI applications at the application level. Table 5is an overview of what we consider logical efforts. The 'light' instruments are probably most effective if they focus on specific risk characteristics. Other 'heavier' tools like setting standards or process requirements have a broader scope.

Table 5 Tools and risk-enhancing characteristics of AI applications in telecom infrastructure

	Autonomy			Predictability			Damage		Scope	
	Training	Validation	Action	Transparency	Deterministic & Linear	I/O	Action framework	Use	Range	Redundancy & Variation
Information and awareness	✓	✓	✓	✓	✓	✓				
Requiring transparency		✓		✓	✓	✓				
Facilitating risk analysis and mitigation	✓	✓	✓				✓	✓	✓	✓
Setting standards		✓		✓	✓	✓		✓	✓	✓
Setting process requirements	✓	✓	✓				✓	✓	✓	✓

5 Conclusions

In this section we answer the research questions that were central to this study (see §1.2).

5.1 Response to the main question

What are the current and future risks of applying AI in the telecom sector? And how can the Dutch Radiocommunications Agency mitigate these risks?

AI applications have specific characteristics that can pose risks when they are used in telecom infrastructures. The extent of autonomous learning but also the unpredictability, action framework and sphere of influence relating to AI application determine the probability and impact of additional risks.

Alongside the *additional* risks of using AI applications in telecom infrastructures, there are (still) conventional risks relating to information security during the entire lifecycle of an AI application: planning, data collection, training, testing, validation and operations. We also see that AI can add specific value for mitigating risks.

Various AI applications interact with each other, with people, with 'normal' automation and possibly the outside world. It is therefore important to assess how AI is applied in the telecom sector at a *systemic level*. We have to consider the ultimate use of these applications that are based on telecom infrastructures, and the level of service they require from the infrastructure.

The Dutch Radiocommunications Agency has various tools at its disposal to mitigate the risks of AI applications in the telecom sector: information provision and awareness raising, stipulating transparency, facilitating risk analysis and mitigation, developing criteria and setting process requirements. At European level, additional requirements could be activated for peripheral equipment.

As starting point, we recommend using tools at a systemic level. There are specific tools for dealing with certain AI risk factors. Broadly speaking, there will have to be a social debate about the desirable level of telecom infrastructures' services.

This research has adopted a specific definition of AI to study its application in the telecom sector. It is conceivable that the conclusions are also relevant in a broader sense for use in autonomous, self-learning and data-driven applications. In the Dutch Radiocommunications Agency's other application domains, there are potentially similar developments and risks in the field of AI.

5.2 Responses to the subquestions

What does the application of AI look like now in the telecom sector and other sectors that make use of digital connectivity?

In the context of telecom infrastructure, AI involves using algorithms based on deep learning, trained using large amounts of data, to automate tasks that could previously only be performed (properly) by humans.

Nowadays we see that most applications of AI in telecom infrastructures involve the optimization of specific parameters. These are strictly defined applications. It is not always clear if what manufacturers call "AI" actually means using algorithms based on deep learning and trained in large amounts of data. After all, algorithmic optimization has been used in telecom infrastructures for many years.

What developments are envisioned in the coming five years for applying AI in the provisioning and use of digital connectivity?

Looking at the coming five years, we see AI applications becoming more and more advanced. A vision shared by a number of suppliers of telecom equipment is that AI will be able to control entire telecom networks. Although it is questionable whether this can happen (entirely) in five years, their vision is definitely one we can anticipate.

What are the risks regarding availability, authenticity, integrity, trust, transparency and predictability in the various sectors as a result of the current and future use of AI? How do we weigh up the risks to the various interrelated aspects in a risk model for digital connectivity?

AI applications have certain characteristics that lead to additional risks for telecom infrastructures. Based on a risk model, we can assess these risks and the characteristics are related to the following aspects of AI:

- **The extent of autonomous learning and implementation of AI.** If this extent is considerable, the likelihood of risk events increases. A significant parameter is whether the AI application is controlled by people or by rules.
- **The extent of the AI application's predictability.** If the models are non-deterministic or highly non-linear, it is more complicated to assess whether an application will work well in all situations. One influential factor is the type of data used and if it can be manipulated.
- **The AI application's action framework.** If the AI application has a highly limited effect on telecom infrastructures, this restricts the impact of a risk event. An application with a wide operating framework has a potentially greater impact.
- **The AI application's sphere of influence**. An application operating at a central level and controlling a telecom infrastructure is more prone to risk than an application that optimises a specific parameter at a low level.

Considering the risks of AI applications in isolation paints a limited picture of the societal risks (as well as advantages) of implementing AI in telecom infrastructures. At the systemic level, the following factors affect the risks:

- **Interaction between AI applications and other systems.** We also discussed this at application level in paragraph 3.2 as correlation of probability and/or effect.
- **Replacing humans with AI.** Having people carry out tasks involves risks, and these can be higher or lower with an AI application. This study does not chart the risks involved with human activities in telecom infrastructures. The model we presented in 3.2 can be used to assess the risks of substituting with AI in order to inform the decision whether or not to implement a human-replacement AI application.
- **Implement AI applications to mitigate risk.** At a systemic level, AI applications can contribute to lowering the level of risk, for example through faster detection of problems or attacks, and by helping to find causes and solutions.
- **Cyber (in)security of AI applications.** AI applications are of course also subject to cyber threats and associated security risks. These risks may increase, because training AI applications involves bringing together large amounts of (sometimes sensitive) data.

Finally, no weighting is given to the scores nor linkage with the negative consequences: the model is not normative with regard to the acceptable level of risk. The model does identify where and what type of AI application with which characteristics can lead to new risks.

How can the Dutch Radiocommunications Agency as supervisory body and implementing organization mitigate these risks?

The Dutch Radiocommunications Agency has various tools at its disposal to mitigate the risks of AI applications in the telecom sector: providing information and raising awareness, stipulating transparency, facilitating risk analysis and mitigation, setting standards and process requirements. As starting point, we recommend using tools at a systemic level. There are specific tools for dealing with certain AI risk factors. Broadly speaking, there needs to be a social debate about the desirable level of telecom infrastructures' services.

6 References

[1] Horáková, J. (2006). *From Golem to cyborg: a note on the cultural evolution of the concept of robots* [www.ceeol.com] Slovenská Akadémia Vied - Kabinet výskumu sociálnej a biologickej komunikácie. pp. 83-98.

[2] MIT (2007). *Contributions and Impact* [jmc.stanford.edu]

[3] Buchanan, B.G. (2005). *A (Very) Brief History of Artificial Intelligence*

[4] Wright, J., and Vesonder, G. (1990). *Expert systems in telecommunications*

[5] Ng, A., Lee, H., Grosse, R., and Ranganath, R. (2011). *Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks*

[6] Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). *ImageNet Classification with Deep Convolutional Neural Networks* vol. 25,

[7] LeCun, Y., Kavukcuoglu, K., and Farabet, C. (2010). *Convolutional networks and applications in vision*

[8] Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [1810.04805]*

[9] Berner, C., Brockman, G., Chan, B., and Cheung, V. (2019). *Dota 2 with Large Scale Deep Reinforcement Learning* [arxiv.org] OpenAI.

[10]Barrat, J. (2015). *Our Final Invention: Artificial Intelligence and the End of the Human Era*St. Martin'S Griffin.

[11]Sutton, R. (2019). *The Bitter Lesson*

[12]Chollet, F. (2019). *On the Measure of Intelligence [1911.01547]*

[13]Ardila, D., Kiraly, A., Bharadwaj, S., and al., e. (2019). *End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography.*Nature.

[14]Arnold, M., Bellamy, R., Hind, M., Houde, S., Mehta, S., Mojsilovic, A., Nair, R., Natesan Ramamurthy, K., Olteanu, A., Piorkowski, D., Reimer, D., Richards, J., Tsay, J., and Varshney, K. (2019). *FactSheets: Increasing Trust in AI Services through Supplier's Declarations of Conformity* [arxiv.org] vol. 63, pp. 6:1-6:13.

[15]Seth, Y. (2019). *BERT Explained – A list of Frequently Asked Questions* [yashuseth.blog]

[16]Zając, Z. (2016). *Bayesian machine learning* [fastml.com]

[17]Hochreiter, S., and Schmidhuber, J. (1997). *Long Short-Term Memory* [doi.org] vol. 9, MIT. pp. 1735-1780.

[18]Haeri, S., and Trajkovic, L. (2017). *Virtual Network Embedding via Monte Carlo Tree Search* pp. 1-12.

[19]de Vries, M., and Albers, J. (2019). *AI in het onderwijs: wat mag er?*Utrecht: Dialogic. pp. 31-46.

[20]Goodfellow, I., Shlens, J., and Szegedy, C. (2015). *Explaining and Harnessing Adversarial Examples [1412.6572]*

[21]Qiu, S., Zhou, S., Liu, Q., and Wu, C. (2019). *Review of Artificial Intelligence Adversarial Attack and Defense Technologies* [www.researchgate.net] vol. 9, p. 909.

[22]Chen, L., Ye, Y., and Bourlai, T. (2017). *Adversarial Machine Learning in Malware Detection: Arms Race between Evasion Attack and Defense*Athene, pp. 99-106.

[23]Finlayson, S.G., Chung, H.W., Kohane, I.S., and Beam, A.L. (2019). *Adversarial Attacks Against Medical Deep Learning Systems* [arxiv.org]

[24]Light Reading (2019). *Guavus Takes Jio's Big Data Challenge* [www.lightreading.com]

[25]Csáji, B.C. (2001). *Approximation with Artificial Neural Networks*Hungary: Eötvös Loránd University.

[26]Naderializadeh, N., Sydir, J., Simsek, M., Nikopour, H., and Talwar, S. (2019). *When Multiple Agents Learn to Schedule: A Distributed Radio Resource Management Framework* [arxiv.org]

[27]DARPA (2020). *Spectrum Collaboration Challenge* [www.spectrumcollaborationchallenge.com]

[28]Dialogic, Radicand Economics & iMinds (2016). *The impact of network virtualisation on the Dutch telecommunications ecosystem: An exploratory study* [www.dialogic.nl] Utrecht: Dialogic.

[29]ITU-T. Focus group on Machine Learning for Future Networks including 5G(FG-ML5G) (2019). *FG-ML5G-ARC5G. Unified architecture for machine learning in 5G and future networks* [www.itu.int] Geneve: ITU-T.

[30]Vriezekolk, E. (2016). *Assessing Telecommunications Service Availability Risks for Crisis Organisations* [research.utwente.nl] Universiteit Twente.

[31]Kinney, G., and Wiruth, A. (1976). *Practical Risk Analysis For Safety Management (No. NWC-TP-5865)*China Lake, CA: Naval Weapons Center.

[32]NCTV. *Overzicht vitale processen* [www.nctv.nl]

[33]RTL Nieuws (2020). *Agent aangevallen in Tilburg, noodknop werkt niet: 'Dit moet echt opgelost worden'* [www.rtlnieuws.nl]

[34]Sue, J., Brand, P., Brendel, J., Hasholzner, R., Falk, J., and Teich, J. (2018). *A predictive dynamic power management for LTE-Advanced mobile devices*Barcelona, pp. 1-6.

[35]King, L. H. (2019). *This startup uses battery life to determine credit scores* [money.cnn.com]

[36]Kassa, B. (2016). *Quality of Service. Priority and Preemption* [www.npstc.org]

[37]Berghoff, C., Neu, M., and von Twickel, A. (2020). *Vulnerabilities of Connectionist AI Applications: Evaluation and Defence* [arxiv.org]

[38]Parasuraman, R. (1986). *Vigilance, Monitoring and Search*New York: Wiley. pp. 41-1 - 41-49.

[39]Ribeiro, M.T., Singh, S., and Guestrin, C. (2016). *"Why Should I Trust You?": Explaining the Predictions of Any Classifier* [arxiv.org]

[40]G. Fidel, R.B. A. S. (2019). *When Explainability Meets Adversarial Learning: Detecting Adversarial Examples using SHAP Signatures*

[41]The Next Web (2020). *Some Teslas have been tricked into speeding by tape stuck on road signs* [thenextweb.com]

[42]Business Insider (2016). *Here's what actually caused the 2010 "Flash Crash"* [www.businessinsider.com]

[43]Europese Unie (2014). *Richtlijn 2014/53/EU van het Europees parlement en de Raad van 16 april 2014 betreffende de harmonisatie van de wetgevingen van de lidstaten inzake het op de markt aanbieden van radioapparatuur en tot intrekking van Richtlijn 1999/5/EG* [eur-lex.europa.eu]

[44]DARE!! Measurements. *Radio Equipment Directive (RED)* [www.dare.nl]

[45]Agentschap Telecom (2020). *Telekwetsbaarheid* [www.agentschaptelecom.nl]

[46]ETSI. *Industry specification group (ISG) securing artificial intelligence (SAI)* [www.etsi.org]

[47]Korf, R.E. (1997). *Does Deep-Blue use AI?*

[48]Fidel, G., Bitton, R., and Shabtai, A. (2019). *When Explainability Meets Adversarial Learning: Detecting Adversarial Examples using SHAP Signatures*

Annex 1 List of interview participants

Name	Role	Organization
Anne van Otterlo	Account CTO	Nokia
Gerwin Franken	Senior Regulatory Officer	Nokia
Patrick Blankers	Regulatory Officer	Ericsson
Jeroen Buijs	CTO	Ericsson
Jurjen Veldhuizen	Solutions Director	Huawei
Corine van Pinksteren	Regulatory Officer	KPN
Sacha van der Wijer	Head of Advanced Analytics	KPN
Chris Molanu	Lead AI	KPN
Winifred Andriessen	Director Advanced Analytics	KPN
Simone Van Ginhoven	Regulatory Officer	VodafoneZiggo
Aziz Mohammadi	Director Advanced Analytics	VodafoneZiggo
Michiel van Rijthoven	Lead data scientist	VodafoneZiggo
Frank van Berkel	Senior regulatory affairs counsel	T-Mobile
Miruna Anastasoaie	Lead AI	T-Mobile
Steven Latré	Professor Computational Science & Artifical Intelligence	University of Antwerp

Contact:
Dialogic innovation & interaction
Hooghiemstraplein 33-36
3514 AX Utrecht
The Netherlands
Tel. +31 (0)30 215 05 80
www.dialogic.nl/en