HOW TO PROPERLY APPLY SYSTEMS OF ARTIFICIAL INTELLIGENCE

H. Schäbe[1] ,I.B. Shubinsky[2]

[1] Dr.rer.nat., TÜV Rheinland InterTraffic, Cologne, Germany, dr.hendrik.schaebe@gmail.com
[2] DSc, prof., NIIAS, Moscow, Russia, Igor-shubinsky@yandex.ru

Abstract

The authors present their views on the essence of systems with artificial intelligence and point out the limitations to the use of those systems. Based on these considerations, an approach for the correct and effective use of artificial intelligence is proposed. A system with artificial intelligence (SAI) is in fact a very flexible statistical model with many parameters, which cannot be interpreted. Therefore, the use of an SAI is like a brute force attack using a very flexible statistical model to a problem. The sample which is used to train the SAI becomes much more important than the method itself. SAI can be used for safety applications, but the result of an SAI must be verified and that a proof of safety must be maintained. Mostly, this proof must be based on statistical arguments. A best approach for a use of a SAI is if it supports the developer for specific and well specified problem.

Keywords: artificial intelligence, effective application

I. Introduction

Systems using Artificial Intelligence (SAI) are widely used, one could even say that they have become fashionable see e.g. [1-5]. Apart from the obvious and recognised applications, it sometimes seems as if almost all problems could be solved using SAI.
In this paper, the authors present their views on the essence of SAI and point out the limitations to the use of SAI. Based on these considerations, an approach for the correct and effective use of SAI is proposed.

Chapter two presents our ideas about the nature of SAIs. The third chapter discusses the application of SAIs to security systems. The last chapter summarises and discusses the possibilities of effective application of SAIs.

II. The essence of artificial intelligence systems

Many types of SAIs are known, e.g.

- neural networks,
- deep learning,
- machine learning,
- support vector machines.

Vapnik's works [6,7] describe the mathematical core of artificial intelligence systems. In principle, AI systems are very flexible models of mathematical statistics having a huge number of parameters. Moreover, direct interpretation of these parameters is difficult, if not impossible.

This will mean that the data set that is used to train the SAI - in essence it must be a representative sample - has enormous value. In fact, the data almost entirely determines the behaviour of the SAI.

In this way, some problems of mathematical statistics affect the behaviour of the SAI.

The **first problem** is the representativeness of the sample that is used to train the SAI. The sample should cover all cases and situations that are relevant to the problem to be solved by the SAI. The crux of the issue is what is essential to the problem and hence should be reflected in the sample. Since there is no model or other abstract representation - this is the reason for using an SAI - it is difficult to decide which part of the sample is essential and which is not. In the end, it is often necessary to select the sample directly, make it as large as possible and accompany the sampling process with a rough analysis of what factors should be taken into account.

The **second problem** is the verification of the SAI. This means that for statistical verification, a second sample must be available that is completely diverse from the first sample used for training the SAI, but it must also be representative. This can be difficult if there are only single items for special cases that need to be included in both samples, for training and for verification.

The third problem is the difference between model fit to the data on one side versus model predictability on the other. This problem has been known for a long time, cf. [7]. It consists in the fact that a very flexible model with many parameters can approximate the data well, but it cannot predict new, additional cases. There are a lot of models describing well the weather in the past but that have poor predictability of future weather, despite the large number of parameters. SAIs are complex and comprehensive models with a large number of parameters and a researcher using an SAI can fall into this trap - the applied SAI describes the data well but gives a poor prediction for future cases.

The **fourth problem** is false (misleading) correlation. This problem has also been known for a long time [7]. If we study many factors that are not correlated and perform a correlation test with a statistical confidence of 90%, for example, a false correlation may occur with a probability of 10%. Examination of 10 uncorrelated values will thus result in an average of one false correlation. In NRIs, possibly hundreds of correlations are defined as intermediate parameters, e.g. in neural networks or deep learning systems. Thus, false correlations can occur in these SAI parameters and remain undetected.

The **fifth problem** is how to construct a model as a law of nature. There is a notion especially in the social sciences that a law of nature can be derived simply and directly from data using a mathematical formula that approximates the data well. However, it is additionally necessary to have an idea of what this formula expresses and what theoretical imaginations underlie the law. As an example, we can refer to the fact that Einstein's theory of gravity did not arise simply as a generalisation of experimental data, but on the basis of axioms that generalise human knowledge about gravity. Experiments then confirmed Einstein's theory. This process is absent when SAI is applied to process data with the hope of extracting a new law of nature.

In terms of mathematical statistics, SAIs are very complex and flexible models. They need representative samples, which in part require a huge amount of manual work. In addition, at least two different, representative samples are needed.

In this connection, one important concept should be discussed - the concept of **sufficient statistics** [9]. The essence of this concept is that all information contained in a sample of data that fit a parametric model, is contained in these sufficient statistics. Usually these are the parameters of that model. As an example, consider a sample taken from a population of normally distributed numbers. Then all information is contained in the sample mean and standard deviation. Of course, one could apply SAI to such a sample, but then the advantage of use of that parametric model would be lost.

It follows that the existence of a particular model contains additional information that can be used in data processing. This additional information is what makes this processing more efficient. Applying SAI directly does not lead to these results.

To summarise, SAIs are complex and flexible models that describe data well. But they are not able to replace the abstraction process leading to simpler models if an SAI is applied directly.

Of course, SAI can also be applied to the task "Find me a suitable parametric model for the data set" instead of the task "Describe me the data set". However, one should then check the result of the SAI with information about what processes led to the data under study and check the plausibility of the result suggested by SAI.

III Use of SAI for safety functions

An important question is whether the SAI can be applied to safety functions. This task has already been considered in [1]. The answer is in the affirmative. But the application of SAI to safety functions requires certain caveats.

The **first question** is what safety integrity level should be applied and which safety requirements should be derived. Without exception, all functional safety standards prescribe a risk analysis phase in the system life cycle. In this phase, the hazards, the possible accidents caused by them and the corresponding risks are analysed. In this phase, the system is treated as a black box, without yet setting specific requirements for the system development itself. What is derived in this phase are the requirements for the safety functions and methods for reducing the risk to an acceptable level, including the permissible hazardous failure rates of the safety functions of the system, and hence the levels of safety completeness. Since the risk analysis is not based on the internal structure of the future system, it is also acceptable to the SAI. Details can be found in [1].

The development of SAI is partly the same as for other electronic safety systems: hardware and software are developed according to the requirements of standards and the safety integrity level.

But, such a system is not yet operable, it needs to be trained with a large and representative sample. Thus, in addition to the classical elements of hardware and software, a third element is included, and that is the training sample.

The second question is about this third element. It also complicates the proof of safety, because one has to take this third element into account.

If we follow the ideas of [2], there are two approaches.

The **first approach** is based on the statistical ideas of the proven in use approach. In this case, the SAI is tested using a second, independent, representative sample. This sample should be of sufficient size to allow statistical proof for the required level of safety integrity. Especially for the third and fourth safety integrity levels, sample sizes that are needed that are practically difficult to obtain. Note that a second sample is required for all SAIs for system validation reasons. However, more stringent requirements apply for safety proofs, since this sample is used for statistical proofs.

The second approach is based on the possibility of explaining the behaviour of SAI. Here we can selectively cite the work of [10]. This standard provides a comparison with the ISO 26262 standard [11]. Further, safety management for artificial intelligence is introduced. The sampling issues for training and verification of SAI are thereby covered. It describes the lifecycle, methods of data verification and validation, conducting relevant safety analyses, the process of training of the SAI, the approach to failures, and so on. This standard is just one example of the development of standards for SAI in various fields.

Approximately in the same manner the argumentation in [12] is carried out. Here it is proposed to prove safety by a combination of formal methods, statistical methods and with the help of explicable SAI. And this paper is only an example of a number of articles.

One important aspect of safety systems is the tolerable hazardous rate. For conventional safety systems without the use of artificial intelligence, the hardware must have a hazardous failure rate less than an acceptable value corresponding to the safety level. Software has only systematic failures, which can be neglected if the software is designed following the requirements of the applicable safety integrity level. The same is true for systematic hardware failures. And they can be neglected if the development followed the requirements of the standard.

For SAI, there is an additional category of dangerous failures caused by erroneous decisions of artificial intelligence algorithms. This problem can be solved in such a way that a part of the permissible intensity of failures refers to erroneous and thus dangerous reactions of artificial intelligence algorithms. Consequently, the SAI for the safety function has the same total dangerous failure rate as a system without the use of artificial intelligence. This is achieved at the cost that the hardware for SAI must be better in order to transfer some of its tolerable hazardous failure rate budget to artificial intelligence.

IV. Conclusions

The question now is how to use SAI effectively. In fact, the use of an SAI is like a brute force attack using a very flexible statistical model to a problem. It must be remembered that the result of an SAI must be verified and that a proof of safety must be maintained when SAIs are used in systems with safety functions.

- **SAI supports the developer**

In this approach, the SAI collects information, compiles a list of references, etc. In this way, the SAI supports the developer, but the developer takes full responsibility for the results of his work.

- **Selection of the mathematical model**

Artificial intelligence helps in the selection of a mathematical model. In this case, the artificial intelligence suggests a model, and the developer checks the applicability of the model and shows with the help of statistical methods that the model is suitable. And in this case, the full responsibility lies with the developer.

- **Data analysis**

The SAI is used to analyse data. Trends, relationships, etc. can be identified. However, it is recommended to look for patterns that led to the data. In this case, the responsibility remains with the developer

These three cases are only selected examples of the application of SAI. It is recommended that the SAI be used as an assistant who carries out routine work, the results of which are checked by the developer himself.

In this way it is possible to avoid voluminous proofs of correctness of the SAI or even proofs of safety if the SAI would be used in systems with safety functions.

In any case, the use of SAI should not lead to the neglect of natural intelligence.

References

[1] Schäbe, H. and J. Braband, J., (2020). On safety assessment of artificial intelligence, *Dependability*, 19: 25-34.

[2] Schäbe, H. and Braband, J. (2022) The application of artificial intelligence in railway technology for safety-relevant applications - opportunities and problems , Signal and Data Communication 114 / no.5: 14-21.

[3] Shubinsky, I.B., Rozenberg E.N. and Schäbe, H. (2024) Methods for ensuring and proving functional safety of automatic train operation systems, RT&A, 19: 360-375.

[4] Posthoff, C. Computers and Artifical Intelligence, Past – Present – Future (In German: Computer und Künstliche Intelligenz Vergangenheit - Gegenwart - Zukunft), Springer Vieweg, 2022

[5] Berghoff, C., Biggio, B., Brummel, E. Danos, V., Doms, T., Ehrich, H., Gantevoort, T. Hammer, B., Iden, J., Jacob, S. Khlaaf, H. Komrowski, L., Kröwing, R., Metzen, J. H., Neu, M., Petsch, F., Poretschkin, M. Samek, W., Schäbe, H., von Twickel, A., Vechev M., and Wiegand, T., Towards Auditable AI Systems Current status and future directions, TÜV Verband, BSI, Fraunhofer, 2021

[6] Vapnik, V.N., Chervonenkis, A.Y., Pattern Recognition Theory. Statistical problems of learning (In Russian). Moscow, Nauka, 1974

[7] Vapnik, V.N., The Nature of Statistical Learning Theory, Springer, 2010.

[8] A.N. Kolmogorov, The Theory of Probabiity and Mathematical Statistics, Moscow, Nauka, On the question of applicability of prediction formulae, derived by statistical methods (in Russian), 161...167

[9] L. Schmetterer Introduction to Mathematical Statistics (In German: Einführung in die

mathematische Statistik,) Springer, Wien, New York 1966, I

[10] Road Vehicles - Safety and artificial intelligence, ISO DPAS 8800, 2024

[11] ISO 26262 *Road vehicles - Functional safety, 2018*

[12] G. Henzal, T. Strobel, J. Großmann, B.-H. Schlingloff, M. Leuschel, S. Sadeghipour, J. Firnkörn, (2021) KI-LOK - A joint test procedure project for AI based components used in railway operations, Signal and Data Communication (113) 10/2021 p. 6-15