Artificial Intelligence Landscape – An Introduction in Technology Fields & Research Areas

Preprint · March 2019

DOI: 10.13140/RG.2.2.31652.53128

4 authors:

Helmut Leopold
AIT Austrian Institute of Technology
97 PUBLICATIONS **330** CITATIONS
SEE PROFILE

Willibald Krenn
AIT Austrian Institute of Technology
31 PUBLICATIONS **233** CITATIONS
SEE PROFILE

Ross King
AIT Austrian Institute of Technology
158 PUBLICATIONS **814** CITATIONS
SEE PROFILE

Cristinel Mateis
AIT Austrian Institute of Technology
34 PUBLICATIONS **609** CITATIONS
SEE PROFILE

Artificial Intelligence Landscape

-

An Introduction in Technology Fields & Research Areas

Helmut Leopold, Willibald Krenn, Ross King, Cristinel Mateis

AIT Austrian Institute of Technology
Center for Digital Safety & Security

Vienna, 17 March 2019

Summary

Following a brief explanation of the terminology associated with Machine Learning and Artificial Intelligence (AI), this document describes a method to identify different AI research and development disciplines in the context of an eco-system. Based on this method, relevant AI aspects are briefly explained, and emphasis is placed on the fact that different research areas must be addressed if AI is ultimately to be deployed in multiple application areas in an effective, economically sound and socially sustainable manner. Last but not least, numerous examples of the successful application of AI in a variety of fields are elucidated.

Content

1 Introduction

Artificial Intelligence (AI) is currently regarded across all economies of the world as one of the most important topics in the area of digitalization for both economical and societal development [1]. As a result, there are corresponding strategies and widespread debates in almost all industrialized countries [2, 3]. This also applies to Austria, where a wide-ranging AI strategy process was initiated in 2019 [4].

In this document, we provide a general overview following a brief description of the development history of AI. We explain the most important terms involved and describe the principal learning methods associated with machine learning. We then elucidate eight important topical areas that must be given due consideration for an AI research strategy and the operational planning of this new technology in an industrial and social context. Finally, we draw up a list of the latest AI application examples from research and industry projects currently underway.

2 Terminology

Due to the availability of high yet inexpensive computing power, which was mainly driven by the global gaming industry market in the wake of the development of GPUs (Graphical Processing Units), artificial intelligence (AI) has recently become a hype topic, with this term being understood and used in a variety of ways [5, 6, 7].

A fundamental debate - as to when a machine is deemed intelligent and how intelligence can be defined - has been underway among computer scientists and philosophers for decades. In the 1950s, a pioneer of information technology, Alan Turing, conceived the so-called Turing Test to determine the intelligence of a machine based on its interactions with people in certain communication situations. Up until now, however, tests have shown that people tend to quickly ascribe intelligent behavior to machines even if such “intelligent” behavior is only simulated through the performance of very simple functions.

The increasing performance of computer systems has also spurred a broad debate on the question whether machines can achieve human intelligence or even outperform and ultimately dominate mankind. This time in history is referred to as “singularity” and described as “super intelligence” [8, 9, 10].

However, if one takes a closer look at the technical development of information technology over the past decades and compares this to the various prediction, it becomes obvious that technological evolution has often given rise to excessive expectations as well as fears and machines simply remain tools for us humans [11].

Regardless of this fundamental discussion and philosophical considerations, AI systems are nowadays often defined as machines whose decision-making processes not only build upon mathematical models or procedural processes such as the conventional program logics of software development, i.e. rule-based systems, but also refer to new forms of learning machines (so-called neural networks – explanatory notes will follow).

For our further considerations, however, we will use a definition from another perspective based on a significantly narrower understanding of this term as summarized below.

2.1 Artificial Intelligence

Systems with artificial intelligence (AI) build upon machine learning methods that are based on approaches that rely on information technology and mathematics as well as on neural networks. Their special problem-solving capabilities can be successfully and effectively applied in a number of very concrete and limited application fields, such as - for instance - the control of robots and industrial plants or for image recognition purposes.

In addition to machine learning methods, AI systems also require other functions to be able to make decisions, deduce forecasts from learned data or find approaches to pursue specific decision-making objectives. Thus, AI describes the deployment of systems capable of learning in certain concrete application scenarios.

2.2 Machine Learning

The term "machine learning" (ML) describes the methods of a learning process; i.e. the learning of a specific function by a machine. This can occur using different approaches such as decision trees, mathematical methods (for instance statistics) or even neural networks (NN). Machine learning is based on methods that search for recurring patterns in data, recognize rules and divide data into classes.

2.3 Deep Learning

AI systems are normally referred to as "Deep Learning Systems" or "Deep Neural Networks" when they are based on neural networks (NN).

2.4 Feature Extraction – Embeddings

In the case of conventional machine learning methods that do not build upon neural networks (NN), it is necessary to determine a set of data values and/or patterns as generalized elements (features). Based on these features, specified data structures can be identified by automated processes. Feature extraction is an important method for recognizing objects or events in data types such as images, videos or audio files.

In the case of neural networks (NN), it is no longer necessary to predetermine the searched data structures in this way as the concept of NN already implies a pattern recognition process, which takes place through the neural network learning process. These features that are learned by the NN are referred to as "embeddings".

3 Machine Learning – Learning Method Types

The basic principle of AI is that AI algorithms are first trained with training data and are then able to make decisions based on real data following a preliminary learning phase. AI learning methods are basically divided into three main classes:

- supervised learning
- unsupervised learning
- reinforcement learning.

When machines are trained with specific training data for a concrete type of problem, which can then be applied to other problem areas, this is also known as "transfer leaning".

One should bear in mind that training data should produce a "learning effect" and should not simply learned "by heart" by a machine. This is referred to as overfitting and can occur when too many free parameters are made available to an AI system. To prevent overfitting, available data sets are divided into training and test data. If the AI system achieves better results with the training data than with the test data, this might be an indication of potential overfitting.

As training data can influence the whole system and have the potential to even manipulate it, the testing, verification and explainability of AI is of fundamental importance (see paragraph 5.2 below).

The following paragraphs summarize the most important concepts.

3.1 Supervised Learning – Ground Truth

In a "supervised learning" scenario, the machine receives feedback about its learning success from the expert while the learning process with the training data is taking place. For this kind of approach, a pre-defined objective for the targeted analysis outcome, a so-called "ground truth" is required. An expert must determine which set of data contributes to achieving the desired outcome and which does not. This can only be determined by a human expert, who has specific knowledge of the system concerned, resulting in important requirements for the generation of training data for the AI algorithms. Thus, a large set of training data is necessary, training data must be appropriate to enable the achievement of the right learning effects and data should not contain any errors as they would otherwise be learned by the AI system (for bias problems in training data see Paragraph 4.4 below).

3.2 Unsupervised Learning

In an "unsupervised learning" scenario, there is no need for a "ground truth" for the training data. The system is required to just discover certain patterns without the need to recognize the meaning of these patterns within a global context. The end-result of unsupervised learning approaches is basically a form of clustering[1].

When unsupervised learning approaches are used, an AI algorithm can detect a system's normal operating mode and, once the system is put into operation, recognize anomalies. In this case, it is of vital importance that training data are made available in such a way that potential anomalies are not included, which often proves to be quite a difficult requirement to fulfil.

3.3 Reinforcement Learning

"Reinforcement learning" is a third fundamental approach in AI research, which - along with supervised and unsupervised learning - provides an effective learning method. This approach consists in fine-tuning a neural network, which has been previously pre-trained on a certain dataset, while using real data once the system has entered the operating mode. In a "reinforcement learning" scenario as opposed to a supervised learning scenario, no pre-defined outcome is learned via training data, but the AI system is trained based on permanent reward and punishment feedbacks: i.e. there is basically no right or wrong answer but a formulated goal and the AI system learns how to achieve this goal through ongoing empirical knowledge. This learning paradigm is often used in robotic systems.

4 AI Technology Landscape

With a view to classifying the relevant AI research fields, we suggest the following eight different thematic areas, which are also presented in figure 1 below:

1. Application fields – markets
2. Data sources
3. AI objectives
4. AI operating conditions (security of methods, testing and verification)

[1] In order to group similar data sets into clusters, the k-means clustering algorithm or the hierarchical clustering method are used.

5. AI training data/ground truth
6. Data science tools (data management and processing tools)
7. AI technologies - basic know-how topics (mathematics, NN architectures, etc.)
8. Hardware platforms

These thematic areas are explained in further detail in the following paragraphs.

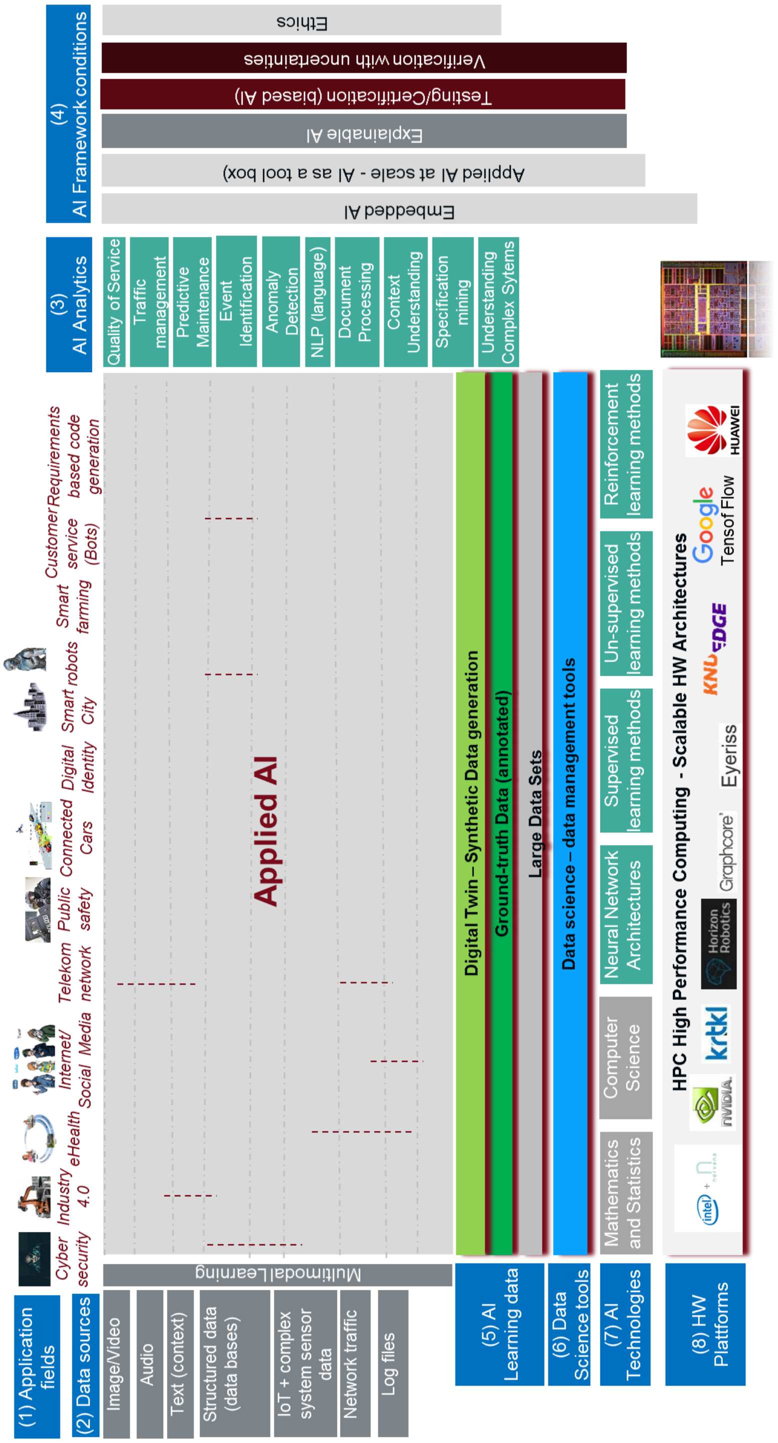

Figure 1: AI technology landscape

4.1 AI Application Fields

AI methods are used in different markets, i.e. application fields. Current application examples for the successful use of AI methods include:

- Cyber security
- Industry 4.0/production
- Medicine
- Internet/social media
- Telecommunications networks
- Public safety
- Connected cars
- Autonomous vehicles
- Digital identity (recognition of biometric features)
- Smart city
- Robots
- Smart farming
- Customer service (bots)
- Requirement-based code generation

In all these application fields, different data sources provide the basis for carrying out analyses and deriving decisions from them.

4.2 Data Sources

The application of AI algorithms takes place on the basis of different data sources. Different data sources require a wide range of methods and tools in order to process data in an appropriate manner for the AI algorithms. Different classes of data include:

- Unstructured data:
  - Images
  - Video
  - Text
  - Audio
    - Environmental
    - Speech
  - IoT and complex systems
  - Network data
  - Log-files of software applications
  - QoS parameters of network protocols (network traffic)
- Structured data:
  - Databases
  - CSV files[2]

4.3 Objectives of an AI Analysis

AI analytical methods can be deployed for a wide range of objectives. Examples of AI approaches include:

[2] CSV (comma-separated values) is a simple file format used to store tabular data, such as a spreadsheet or database.

- Quality of Service (QoS): monitoring and evaluation of quality of service in IT systems
- Traffic management: monitoring and analysis of data streams in networks and communications protocols (generally speaking for Internet und telecommunications network operators)
- Predictive maintenance: analysis of technical parameters to optimize maintenance and operating processes
- Event detection/identification: detection of specific events/patterns in larger data blocks
- Anomaly detection/identification: detection of events/patterns, which deviate from typical system behavior
- NLP - Natural Language Processing: speech-to-text conversion, understanding the context of a language (intent recognition), automated translation
- Automated document processing: recognition of contents in documents. This strongly depends on the comprehension of a language and the document structure.
- Context understanding in applications und software systems
- Specification mining: learning specific system features from the observable behavior of the system in a logically comprehensible way
- Understanding complex systems.

4.4 Training Data

To make sure that supervised AI methods work effectively, a large quantity of training data is needed, with data quality playing an equally important role. However, even if there is a large set of data, given the fact that these datasets might include numerous corrupted data elements – noisy data – or even incorrect data – biased data –effective results will not necessarily be attained.

Thus, a number of different requirements for the generation and use of training data must be fulfilled:

- **Large data sets**: A large amount of appropriate data is needed to train AI algorithms. Massive datasets can be easily collected in certain applications. In a globally accessible search engine or globally deployed translation service on the Internet, a large amount of training data can be produced in a very quick and economic manner. In an Industry 4.0 context, such data collection processes are not easy to realize. For this reason, there is an inherent problem in the industrial field for the development of effective AI algorithms.
- **Digital twin – synthetic data generation**: Since data cannot always be easily obtained from operating processes, special methods are designed to synthetically generate appropriate test data.
- **Annotated ground-truth data**: In a “supervised learning” scenario, the machine receives feedback about its learning success from the expert (person) while the learning process with the training data is taking place. For these approaches, a pre-defined objective for the targeted analysis outcome, a so-called “ground truth”, is required. An expert must determine which set of data contributes to achieving the desired outcome and which does not. This can only be accomplished by a human expert, who has specific knowledge of the system concerned, resulting in important requirements for the generation of training data for the AI algorithms. Large corporations, like Google for instance, constantly publish datasets that are commented with the correct answers (ground truth) to make them available to the research community [13].
- **Data science tools**: effective tools for the elaboration of large data volumes are imperative – see the following paragraph.

4.5 Data Science as a Basis for Effective AI Application

A wide range of data science tools are used for the processing of large data volumes such as data preparation, presorting and other search functions. Open source solutions in this regard include:

- Scalable data storage: Hadoop, Cassandra
- Scalable data processing: Spark
- Programming: R, Scala, Python
- Machine Learning libraries: Weka, scikit-learn, Spark-ML (MLlib), TensorFlow, TensorForce, Theano
- Natural Language Processing (NLP) resources: Stanford-NLP, PyTorch, Flair, NLTK
- Frameworks to use and elaborate neural networks (NN), such as: TensorFlow (Google), Keras (user-friendly TensorFlow), PyTorch (Facebook) or MXNet (Amazon).

To be able to use these data management tools and establish a basis for effective AI algorithms, it is imperative to extrapolate useful and sound datasets from the systems under consideration.

Large datasets alone by no means represent a source of knowledge. First, raw data must be given a meaning within a context with other data to generate information that is going to be processed and interpreted in appropriate applications in such a way as to generate knowledge.

Depending on the case of application, specific information and preliminary work are needed to achieve sound, solid and reliable results from AI algorithms, such as:

- Natural Language Processing (NLP): to achieve good results from AI algorithms specific information is needed for the respective language, such as stop-words, word trees and word embeddings, etc.
- Rule-based systems: these require systematic formalization of expert know-how in a machine-readable form.

Especially in an Industry 4.0 context, the availability of appropriate sensor and machine data is of pivotal importance. In this case, it is necessary that data and domain experts join forces and engage in an intensive exchange of information in order to evaluate data sources, elucidate data generation strategies and define specific data collection strategies. This essential aspect is often underestimated in many AI discussions.

4.6 Neural Networks and Fundamental AI Technologies and Areas of Competence

In order to carry out comprehensive data management tasks, a deep understanding of the different data management tools such as scalable data storage, data processing and programming (see paragraph above) is key. In addition, AI is based on multiple mathematical and information technology disciplines such as linear algebra, statistics (linear, logistic regression), and machine learning (Random Forrest, heuristic methods, probability calculation, trial/error, correlation, etc.).

Neural networks (NN) build upon a wide range of architectures and learning methods and are described on the basis of different parameters. These include, for instance, the number of layers, input neurons, output neurons, feedback loops, layers with or without learnable weights, threshold value calculation with different functions etc. Depending on the design principles used, different architectures and operating modes can be achieved, which are subject to scrutiny in diverse research projects. These include among others Feed Forward Networks, Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Extreme Learning Machine (ELM), Markov chains and Hopfield

Networks. In addition to these types of neural networks, traditional machine learning approaches such as Support Vector Machines (SVMs) are crucial technologies to be considered for machine learning and AI solutions."

In an AI application, different network architectures can be connected with one another. For instance, the neural network AlexNet consists of 25 layers, of which 5 have learnable weights and the last three layers are fully connected [16]. The neural network VGG16, also called OxfordNet, is a convolutional neural network (CNN) featuring 41 layers, of which 16 have learnable weights; 13 are convolutional layers and 3 layers with learnable weights are fully connected [17].

Thanks to evaluation and research projects currently under way, more hands-on experience and in-depth knowledge are being gained on the basis of the observed advantages and drawbacks of the different approaches used. For instance, CNN are proving particularly suitable for image processing tasks, while RNN are most suitable for time-series datasets.

The development of these different neural network architectures along with different learning models for diverse data types, solution objectives and application fields are the current subject of global AI research.

4.7 AI Hardware Platforms

AI methods that are based on CNN in particular have recently achieved a breakthrough thanks to the availability of high-performance processors, which were driven forward by the gaming industry and the related development of GPUs (Graphic Processing Units). In the meantime, various manufacturers have specialized in the development and production of high-performance and highly scalable hardware platforms specifically designed for AI. These include, Nervane, which was acquired by Intel, nvidia, krtkl, horizon robotics, Graphcore, Eyeriss, KNUEdge, etc.

Tensor has been developed by Google as a hardware platform specifically designed for AI applications [18]. In the same way, Huawei has also launched a special AI chipset onto the market [19].

At the same time, concerted efforts are being made at both the European and global level in the context of High Performance Computing (HPC), which are aimed at developing high-performance computers and making their computing power available to research and industry.

4.8 AI Framework Conditions

With a view to promoting the widespread use of AI in numerous application fields, it is of pivotal importance to define further approaches that guarantee an industrially, economically but above all socially acceptable deployment of AI. Such important focus areas are:

- **Applied AI at scale - AI as a tool box**: AI methods should be generically applicable and not specifically designed for a particular application field. Thus, the search for learning models that are suitable in practice, i.e. NN architectures and data analysis models pose fundamental challenges. Besides, scalable and economically sound approaches need to be identified to annotate ground truth in training data appropriately.
- **Embedded AI**: the use of AI methods in hardware-oriented environments with limited technical resources (CPU, storage space, etc.). This is particularly important for industrial application scenarios.
- **Explainable AI:** AI methods must be understandable and explainable by humans despite their inherent lack of transparency. In all areas, where control devices are concerned, and the reliable functioning of a system is deemed a basic requirement (safety), the use of AI is not possible as one cannot transparently understand why AI algorithms lead to certain

outcomes. Therefore, new methods are needed to make the result-finding process of AI algorithms explainable and more transparent in order to be able to use the latter in deterministic systems [12]. For this purpose, dedicated test methods for AI algorithms and verification approaches are necessary even if no clear relation exists between data sources and analytical results (verification with uncertainties). See also Paragraphs 5.1 and 5.2 below.

- **Ethics**: AI methods and analytical results must be subordinated to basic ethical principles and be verifiable through appropriate test methods. Research examples show how AI algorithms tend to take unethical decision paths if unsuitable training data are used. This means that approaches are needed that guarantee the prevention of any kind of discrimination and social injustice driven by AI algorithms. "Explainable AI" along with suitable test methods are therefore vital approaches and important pillars to ensure ethical principles in AI systems [14, 15].

  The basic social dimension of the current trend towards increasing automation in numerous working areas represents a further important issue that requires an intensive public debate in society at large (see Paragraph 7 below).

All these topics are new thematic areas and are the subject of new research fields.

5 New Research Fields

5.1 Explainable AI

System builders should always have asked themselves a few questions with regard to system design and implementing complex decision-making systems, such as: "Is the system working as intended?", "Do the decisions being made seem sensible?" or "Are we conforming to legislation as well as rules of ethics such as equality regulation?". These questions also give rise to other concerns such as "What should I do differently to get another system decision next time? These issues have always been relevant, but with the growth of AI and machine learning algorithms they have acquired a new significance (Bodo et al., 2017; Kroll et al., 2016; Nissenbaum, 1996; Olhede and Wolfe, 2018; Pasquale, 2015; Selbst and Barocas, 2018; Veale and Edwards, 2018) [12].

AI that builds upon neural networks further exacerbates these issues as black box functions are applied. Decisions are derived from machines that use millions of independent values, which defy human comprehension.

New approaches are needed to be able to make reliable predictions about the areas in which AI machines are likely to make decisions. But, of even greater importance is the question as to under which circumstances decision-making machines are going to fail. Following the emergence of AI, the discipline known as explanation sciences, which ranges from philosophy up to concrete technical test systems, is bound to face a completely new set of challenges moving forward.

5.2 When do AI Systems Fail? - Testing and Verification of AI Systems

Questions as to how AI systems are tested for compliance with specific requirements, and more concretely, how it can be established when AI systems fail, and no sensible outcome is achieved, have given rise to a new discipline, which will present us with considerable challenges moving forward.

Following the introduction of AI systems and the necessity of testing and verifying them, a new challenging topical area has emerged. Compared to conventional systems, verification tasks are proving even more challenging especially due to the self-learning feature of AI. Since a system or a function is learned and, therefore, no construction manual is available, the pivotal question is what ultimately "correct system behavior" means? Besides, systems that can adapt to new environmental conditions during runtime are inherently more complex than systems with a pre-defined static behavior. Thus, it goes without saying that explainable AI models facilitate the verification problem.

An example that demonstrates the vulnerability to attacks of learned functions is provided by the paper published by Sharif et al. on the vulnerability of face recognition algorithms to color-coded eyeglass frames [20]. Similar examples of adversarial AI exist also for traffic sign recognition systems [21], resulting in massive repercussions for the safety and security of the entire system in an automated driving context.

Data can be altered in such a way that the perturbation of the original dataset is almost imperceptible, but AI algorithms are completely fooled as demonstrated by an example of a picture with a slight change in pixel. In this case, the AI algorithm responsible for analyzing the picture of a bus was fooled into thinking that the depicted motif was an ostrich [22]. Another problem regards the fact that imperceptible and irrelevant data can be learned that have no importance for the final outcome of the analysis [23]. Against this backdrop, the issue of machines potentially becoming increasingly "superstitious" takes on a new significance. The behavioral scientist Skinner was able to recognize this issue as early as 1948 based on his experiments with pigeons [24].

Google has also addressed this fundamental issue by organizing the "Unrestricted Adversarial Examples Challenge" [25, 26, 27]. Following this initiative, the international research community is invited to identify such problem areas, allowing Google to correspondingly improve its AI algorithms and protect them against such attack scenarios.

Other current approaches relating to the verification of AI systems combine the automatic creation of test inputs with a quantitative analysis of the resulting system behavior with regard to predefined limits that cannot be exceeded. In a second step, attempts are made with heuristic methods to alter the inputs over multiple interactions in such a way that the output of the system is very close to one of the predefined limits. The ultimate objective is to make the system exceed this limit and therefore demonstrate the faultiness of the system.

For the time being, there is no established and recognized approach for dealing with the verification challenges posed by AI-based systems. Especially with regard to the integration of AI algorithms in the safety context, there is still a substantial need for research. For instance, it is not possible to use AI algorithms in current automotive industry standards beyond the SIL2+level[3].

6 Application Examples of Current AI Systems

6.1 AI Application Examples in the Area of Infrastructure and Security

Examples of successful applications of AI systems include:

- Monitoring and evaluation of QoS – Quality of Service (in communications networks – BigDama project https://bigdama.ait.ac.at/ [28]

[3] Automotive Safety Integrity Level (SIL) is a risk classification scheme defined by the ISO 26262 - Functional Safety for Road Vehicles standard. ISO 26262 is an international standard for the functional safety of electrical and/or electronic systems for the production of automobiles.

- Detection of events/patterns, which deviate from typical system behavior (anomaly detection), in log data of IT applications to identify potential cyber-attacks; AECID; https://aecid.ait.ac.at/ [30]
- Image and audio analysis of massive data volumes for the fight against crime and terrorism – EU project VICTORIA; https://www.victoria-project.eu/ [31, 32]
- Analysis of large amounts of text such as court records and police files – EU Project COPKIT; https://copkit.eu/ [33]
- NLP Natural Language Processing: understanding the context of a language (intent recognition) for a language user interface to operate the document management system DOXIS of the SER company [29].

6.2 AI Application Examples in the Area of Marketing and Sales

Additional typical examples of AI applications in customer services in the marketing, sales and commercial context include:

- Recognition of contents in documents. This strongly depends on the comprehension of a language and the document structure (automated document processing).
- Automatic machines to process customer queries and perform customer service tasks (Chatbots)
- Analysis of customers' purchasing behavior
- Optimization of commercial offers based on the predictions made with regard to customers' purchasing behavior
- Optimization of inventory management in supply chains; prediction of potential demand for specific stocks and evaluation of risks with regard to the planned delivery periods.

6.3 AI Application Examples in Industry

Current examples of AI applications in the industry include:

- Predictive maintenance: analysis of technical parameters in order to optimize maintenance and operating processes and prevent business interruptions [34]
- Learning specific system features from the observable behavior of the system in a logically comprehensible way (specification mining); to predict, for instance, certain system behaviors such as the response time of web services [35]
- Provision of effective test data. Use of AI to create effective test cases and test data for the testing of digital systems [36]
- AI for signal processing for new sensor functions [37]

7 The Social Dimension of AI

Since AI systems are well-suited to automate simple workflows, they also pose special challenges in the social context. Regardless of whether they involve more effective robots in factory halls or specific software systems that lead to service automation in our information society, AI is introducing a new dynamic into the work automation issue. Against this backdrop, the need for further education and professional training measures along with upskilling on the labor market acquires new significance, which we all need to address with appropriate answers.

8 Literature

[1] John P. Holdren, Megan Smith, *Preparing for the Future of Artificial Intelligence*, Obama Whitehouse, Executive Office of the President National Science and Technology Council Committee on Technology, October 12th, 2016.

[2] Joshua New, *Why the United States Needs a National Artificial Intelligence Strategy and What It Should Look Like*, December 4th, 2018, http://www2.datainnovation.org/2018-national-ai-strategy.pdf

[3] Gregory C. Allen, *Understanding China's AI Strategy - Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security*, February 6th, 2019, https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy

[4] *AI Strategie Austria*, BMVIT Bundesministerium für Verkehr Innovation und Technologie und BMDW Bundesministerium für Digitalisierung und Wirtschaft, https://www.bmdw.gv.at/DigitalisierungundEGovernment/Documents/AIM_2030.pdf

[5] Hannes Androsch et al. (eds.), *TECHNOLOGIE IM GESPRÄCH: KÜNSTLICHE INTELLIGENZ*, Jahrbuch zu den Alpbacher Technologiegesprächen 2018 (ISBN: 978-3-903207-24-0), https://buch.verlagholzhausen.at/current/article/technologie-im-gespraech-kuenstliche-intelligenz-jahrbuch-zu-den-alpbacher-technologiegespraechen-201-1/

[6] Stuart Russel, Peter Norvig, *Künstliche Intelligenz – Ein moderner Ansatz*, 3. überarbeitete Auflage, Teil VII Schlussfolgerungen, Philosophische Grundlagen, S. 1195-1196; Pearson Deutschland GmbH, München, 2012.

[7] *Artificial Intelligence, Robotics and 'Autonomous' Systems*, European Group on Ethics in Science and New Technologies, European Commission, Directorate-General for Research and Innovation, Scientific Advice Mechanism, March 2018.

[8] BBC News, *Stephen Hawking warns artificial intelligence could end mankind*, 2 December 2014, https://www.bbc.com/news/technology-30290540

[9] Stephan Bader, *KI – Wenn der Fortschritt plötzlich explodiert*, Open Insights Magazin, 13. Januar 2018, https://www.openinsights.de/kuenstliche-intelligenz-wenn-der-fortschritt-ploetzlich-explodiert/

[10] Stefan Bornemann, *Die künstliche Super-Intelligenz – das Risiko Künstlicher Intelligenz*, lead&conduct!, 27. Dezember 2017, https://lead-conduct.de/2017/12/27/die-kuenstliche-super-intelligenz-das-risiko-kuenstlicher-intelligenz/

[11] Christof Baumgartner, *KI fehlt es an Intelligenz*, Computerwelt, 12. September 2018, https://computerwelt.at/printausgabe/ki-fehlt-es-an-intelligenz/

[12] Brent Mittelstadt, Chris Russel, Sandra Wachter, *Explaining Explanations in AI*, Proceedings of FAT* '19: Conference on Fairness, Accountability, and Transparency (FAT* '19), January 29–31, 2019, Atlanta, GA, USA. ACM, New York, NY, USA, doi/10.1145/3287560.3287574, ISBN: 978-1-4503-6125-5 (University of Oxford - Oxford Internet Institute, University of Surrey).

[13] https://ai.google/tools/datasets/

[14] H. Leopold, *Beherrschen wir die künstliche Intelligenz?*, in: OVE GIT Newsletter Social Media, November 2018, http://archive.newsletter2go.com/?n2g=884ei2cz-mb6e4z8y-10az

[15] A. Schindler, *Ethics and Bias in Artificial Intelligence*, in: OVE GIT Newsletter Social Media, November 2018, http://archive.newsletter2go.com/?n2g=884ei2cz-mb6e4z8y-10az

[16] Wikipedia, Alexnet, https://en.wikipedia.org/wiki/AlexNet

[17] Francois Chollet, *How convolutional neural networks see the world*, The Keras Blog, 30.1.2016, https://blog.keras.io/how-convolutional-neural-networks-see-the-world.html

[18] Sebastian Moss, *Google unveils second generation TPU, available as a service*, DCD, May 18th 2017, https://www.datacenterdynamics.com/news/google-unveils-second-generation-tpu-available-as-a-service/

[19] *Huawei unveils Atlas artificial intelligence hardware*, MyBroadband Magazine, 16.8.2018, https://mybroadband.co.za/news/hardware/280115-huawei-unveils-atlas-artificial-intelligence-hardware.html

[20] Mahmood Sharif, Sruti Bhagavatula, Lujo Bauer, and Michael K. Reiter, *Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition*. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16), 2016, ACM, New York, NY, USA, 1528-1540. DOI: https://doi.org/10.1145/2976749.2978392

[21] K. Eykholt et al., *Robust Physical-World Attacks on Deep Learning Visual Classification*, IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 2018, pp. 1625-1634, (DOI: 10.1109/CVPR.2018.00175).

[22] Dave Gershgorn, *Fooling The Machine - The Byzantine science of deceiving artificial intelligence*, Popular Science, March 30th, 2016, https://www.popsci.com/byzantine-science-deceiving-artificial-intelligence

[23] Marco Tulio Ribeiro et al, *"Why Should I Trust You?" Explaining the Predictions of Any Classifier*, August 2016, ACM, KDD 2016 San Francisco, CA, USA.

[24] *Die Macht des Aberglaubens, Skinner und seine abergläubischen Tauben*, Spiegel Online, http://www.spiegel.de/wissenschaft/mensch/psychologie-die-macht-des-aberglaubens-a-290219.html

[25] https://github.com/google/unrestricted-adversarial-examples

[26] https://ai.googleblog.com/2018/09/introducing-unrestricted-adversarial.html

[27] Brown, T.~B., Carlini, N., Zhang, C., Olsson, C., Christiano, P., Goodfellow, I., *Unrestricted Adversarial Examples*, Cornell University, 2018 (arXiv e-prints arXiv:1809.08352).

[28] Bigdama Projekt, AIT Austrian Institute of Technology, https://bigdama.ait.ac.at/

[29] *Forschungsprojekt „Natural Language Search" erfolgreich abgeschlossen*, F&E Kooperation SER und AIT, AIT Austrian Institute of Technology, 7.3.2018, https://www.ait.ac.at/news-events/single-view/detail/5250/?no_cache=1

[30] *ÆCID - Automatic Event Correlation for Incident Detection*, AIT Austrian Institute of Technology 2019, https://aecid.ait.ac.at/

[31] *EU Projekt Victoria*, AIT Austrian Institute of Technology, https://www.victoria-project.eu/

[32] *Kiras Projekt Florida*, AIT Austrian Institute of Technology, https://www.kiras.at/gefoerderte-projekte/detail/d/florida/

[33] *EU Projekt COPKIT*, https://copkit.eu/

[34] Anahid Jalali et al., *Predicting Time-to-Failure of Plasma Etching Equipment using Machine Learning*, 2019 (submitted to the 2019 IEEE International Conference on Prognostics and Health Management).

[35] R. Schumi, R. Korošec, R. Schlick, B. K. Aichernig, S. Kann, W. Krenn, E. Jöbstl, P. Bauerstätter, and C. Mateis, *Learning and statistical model checking of system response times*, Softw. Qual. J., 2019.

[36] T. Ferrère, D. Nickovic, and A. Donzé, *Interface-Aware Signal Temporal Logic*, 2019.

[37] Florian Wenig, Peter Klanatsky, Christian Heschl, Cristinel Mateis, Nickovic Dejan, *Exponential pattern recognition for deriving air change rates from CO2 data*, ISIE, 2017, pp. 1507-1512.