On Artificial Intelligence: Software and Statistical Issues

Manfred Sneps-Sneppe
Ventspils University of Applied Sciences
Venspils, Latvia
manfreds.sneps@gmail.com

Dmitry Namiot
Lomonosov Moscow State University
Moscow, Russia
dnamiot@gmail.com

***Abstract*—This article is a kind of philosophical essay, a reflection on the difficulties that arise when looking at applications of artificial intelligence (AI) from traditional statistical data processing. In addition, it is associated with an unprecedented amount of Big Data, including the grandiose amount of software. In turn, it raises cyber security issues and requires a new approach to the new AI system auditability, requiring an answer on which statistical indicators to base AI auditability. When we discuss AI applications, it is important to distinguish between autonomy and automation, that is, whether a system is truly autonomous or merely automated. At first glance, it seems that the reason that causes of Big Data analysis failures is the difference in cultures between machine learning and statistical communities. But the reason is apparently deeper, as the statistical paradox in the Big Data example shows. At present, it is not clear whether it will be possible to invent parameters that will help meet the requirements of insurance companies for safety- and security-critical AI applications. It is possible that the two new concepts of Data Defect correlation and the Law of Large Populations discussed in the paper can serve as the starting point of the search for new measures for Big Data. We cannot remain silent about the cyber threat situation either, which makes Big Data analysis extremely difficult. The task of providing robustness of machine learning software, especially in safety- and security-critical areas, is currently beyond the competence of individual companies and even governments and is becoming a problem of international cooperation.**

I. Introduction

This article is a kind of philosophical essay, a reflection on the difficulties that arise when looking at applications of artificial intelligence (AI) from traditional statistical data processing. In addition, it is associated with an unprecedented amount of Big Data, including the grandiose amount of software. In turn, it raises cyber security issues and requires a new approach to the new AI system auditability, requiring an answer on which statistical indicators to base AI auditability.

The AI approach is largely about software and statistics. The rest of the paper is the following. In Section II, we discuss AI (machine learning) limitations. In Section III, we refer to the US Government Accounting Office (GAO) report on cyber threats (2018) and some attempts of the fighter F-35 to turn to AI. Section IV is devoted to the GAO approach how to estimate the status of AI for Weapon Systems. AI Auditability as a very hard task is considered in Section V. Sections VI and VII to discuss the robustness of machine learning software and the gap between machine learning and statistical communities.

II. Artificial Intelligence from the RAND's Viewpoint

Artificial Intelligence, as an academic discipline, appeared in the mid-1950s of the last century. AI refers to computer systems oriented to replicate some human functions and continually get better at their assigned tasks.

Taking into an account the potential magnitude of Artificial Intelligence's impact on the whole of society, and the urgency of this emerging technology international race, President Trump signed an executive order that was designed to ensure (secure) US leadership in artificial intelligence technologies [1]. This is the so-called American AI Initiative, aimed at maintaining American leadership in competition (economic, geopolitical, etc.) with China. This was immediately followed by the release of DoD's first-ever AI strategy [2]. AI has recently become a focus of governments worldwide [3].

For more details on AI military applications, let's refer to the RAND paper [4]. When people want a task done, they either do it themselves or delegate it to another entity, which can be a human or a machine. By delegating, they relinquish some control over how it is done, and the unit performing the task has some autonomy. If a task is perfectly scripted with a defined and known set of rules, technologists say that the unit performing it has "low autonomy" and describe it as "automated". When the unit performing the task is empowered to act without rules or boundaries, it is described as fully 'autonomous'. Almost all tasks performed by machines fall somewhere between these two extremes, so it is useful to discuss AI applications in degrees or levels of autonomy, especially with regard to lethal autonomous weapons systems. It is useful to classify these technologies in a graphical taxonomy that illustrates their interrelationships.

Figure 1 presents such a taxonomy at three levels. Early approaches to AI involved the development of automated systems with the ability to perform scripted tasks according to specific sets of rules. Such approaches are still used to some extent, but in the last couple of decades more sophisticated systems capable of machine learning (ML) have been developed. These systems can gradually improve their performance by recognizing patterns in large amounts of data and taking corrective actions to improve their ability to classify future patterns when they are not specifically

programmed to do so. An even more sophisticated class of ML systems demonstrates deep learning. They use multi-layered artificial neural networks to recognize patterns in data representations, such as labeled images, rather than using task-specific algorithms as in basic ML systems. Recent advances in deep learning using deep neural networks have led to significant improvements in computer vision and image recognition systems.

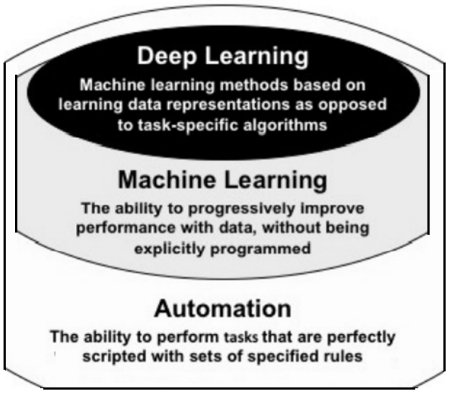

Fig. 1. Taxonomy of AI Technologies [4]

This analysis of Artificial Intelligence relates to quite complicated military applications. Let us name two much simpler cases to illustrate AI Technologies (Fig. 1).

Case 1. Automation: Recruiting. In [5], a matrix for auditing algorithmic decision-making systems is used in the field of recruitment. These screening technologies evaluate applicants in various ways by assessing their suitability for a role, playing online games, analyzing their speech and/or mannerisms to predict performance in the workplace, or analyzing "personality assessment" questionnaires. In this example, the role of AI is clear, because the computer facilitates the work of a part of the staff (perhaps it "weeds out" the talented ones) and the final decision is made by the staff. The next example is much more complicated.

Case 2. Machine learning: plant disease detection. The main point of machine learning is data. To illustrate the amount of data and computer time of machine learning, we use a neural network approach to plant disease detection (Fig. 2). In [6], convolutional neural network (CNN) models were developed to perform plant disease detection and diagnosis using simple images of healthy and diseased plant leaves using deep learning methodologies. Model training was performed using an openly available database of 87,848 photographs. This data includes 25 plant species in 58 different classes of plant and disease combinations, including some healthy plants. The most successful model architecture, the VGG Convolutional Neural Network, achieved a 99.53% success rate in classifying 17,548 previously unseen model plant leaf images (test set). Each image is 256x256 = 65536 pixels. Based on such a high level of performance, it becomes clear that CNNs are very suitable for the automatic detection and diagnosis of plant diseases through the analysis of simple leaf images.

Fig. 2. Leaf images [6]

The total training time for this model was about 5.5 days (!). The learning algorithms were implemented on powerful computers using a parallel programming platform. The classification of a particular unknown image takes an average of about 2 ms. Note that it takes only 2 ms to detect a leaf disease, but before that, colossal highly skilled work was carried out to collect and classify 87,848 photos.

This example raises a difficult question – how to teach botanists? It is enough for the laboratory technician to take a photo of the diseased leaves and identify the disease. But how to match the machine image of a leaf, consisting of 65,536 pixels, with the botanist's representation of dozens or hundreds of macrofeatures of a diseased leaf? Machine learning is unlikely to detect the emergence of a new plant disease. How will an AI algorithm detect it?

III. THE GAO ON CYBER THREATS (2018) AND FIGHTER F-35 TURN TO AI

GAO on cyber threats. In October 2018, the US Government Accounting Office (GAO) sensationally reported [7] that all software-based weapons systems that were tested between 2012 and 2017, including those created over the past ten years, have cyber vulnerabilities and can be hacked. Software updates and limited resources do not allow timely correction of deficiencies. As practice testing showed, programmers already knew about some of the vulnerabilities in weapons systems in advance, since they were identified during previous cybersecurity assessments. For example, one test report states that only 1 out of 20 cyber vulnerabilities identified in a previous test were patched [8]. Is the situation hopeless?

Critics of the GAO [7] seem to have been referring to the Lockheed Martin F-35 Lightning II aircraft (Fig. 3). The F-35 software, which is 8 million lines of software code built into the aircraft, controls most of its functions, including flight control, radar, communications, and weapon targeting. But a large amount of software inevitably has not only errors, but also unpatched vulnerabilities. In addition, the F-35 aircraft works in a network of other aviation and ground systems, which provides additional opportunities for hackers. Any of the connections can be used by enemy cyberwarriors to infiltrate and destroy or disable the aircraft.

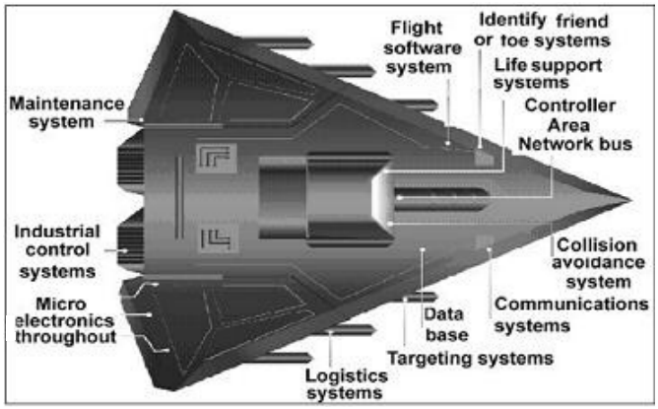

Fig. 3. Embedded software and information technology systems in weapon systems (represented via fictitious weapon system for classification reasons) [7]

The vulnerabilities identified by GAO experts are perceived as a national disaster, because despite more than 20 years and approximately $62.5 billion spent on research and development alone, the F-35 aircraft remains, in all practical and legal senses, nothing more than the very expensive prototype. This leads to the extremely important thought that the original concept of the Joint Strike Fighter was erroneous and went beyond practical technological reality. The U.S. Congress, during the fiscal year 2022 budget debate, refused to authorize additional new orders for the F-35 on top of the Pentagon's already approved requests [31]. What to do? Starting over from the scratch is unrealistic. And how to deal with the existing cyber threats?

On fighter F-35 maintenance tasks. In June 2018, the Joint Artificial Intelligence Center (JAIC) was established. JAIC was a focal point of the DoD AI Strategy [2]. The emergency goal of JAIC was to produce solutions for Predictive Maintenance (the first wave of the DOD's AI strategy, see Fig. 5). The goal was to develop AI-based applications to predict maintenance needs on equipment, such as the E-3 Sentry (known as AWACS, Airborne Warning, and Control System), multirole fighter aircrafts F-16 Fighting Falcon and F-35 Lightning II, as well as Bradley Fighting Vehicle.

According to the recent news [9], the Joint Artificial Intelligence Center will cease to exist come June 1, 2022, as well as two other offices: Defense Digital Service, and Office of Advancing Analytics, or ADVANA. In all three cases, the offices are expected to remain as part of the Chief Digital and Artificial Intelligence Officer (CDAO). CDAO staff will be diffused with the office stovepipes removed.

Relating to AI works, these facts show that the DOD offices are currently in the very beginning of an organizational phase. The AI research is currently in the Expert knowledge stage only, but not in Machine learning (as DOD's experts are estimating), and nothing to talk about the Contextual adaptation stage (see Fig. 5). The DOD's experts estimation of Artificial Intelligence status seems incorrect, a little over-estimated [10].

As an example, we use the F-35 combat aircraft maintenance experience (Fig. 4). Lockheed Martin F-35 Lightning II aircraft is a family of amazing combat aircraft of the future: single-seat, single-engine, stealth, all-weather, multi-purpose, designed for both air supremacy and strike. The aircraft has been developed since 2001. It is assumed that the aircraft will be in service until 2070.

Fig. 4. The F-35 testing by means of ALIS (Lockheed Martin)

Autonomic Logistics Information System (ALIS) is intended to provide the logistics tools for the F-35 program. ALIS consists of several software applications designed to support a variety of squadron activities such as supply chain management, maintenance, training management, and mission planning. During the flight, the aircraft transmits status reporting codes to the ALIS ground station to ensure that maintenance personnel is ready to perform any necessary repairs when the aircraft lands. ALIS, in turn, feeds this data into various databases and engineering models located at Lockheed facilities. It is necessary to check not only the integrity of the data stream, but also the resistance to hacking, cyber espionage, malicious code, etc. The ALIS software contains more than 20 million lines of code.

In 2020, program leaders abandoned efforts to complete the $16.7 billion ALIS system as testing found that no section of the F-35 program was cyber-proof. Pentagon officials have announced that ALIS will be replaced by a new cloud-based system called Operational Data Integrated Network (ODIN) [11]. However, plans have changed. As reported [12] in April 2022, due to multiple factors, including budget cuts, lack of access to proprietary ALIS programming code, and continuous improvement of ALIS, the F-35 Joint Program Office decided to incrementally improve and modernize ALIS instead of replacing it with the new system. US Department of Defense officials renamed the system ODIN (probably to hide miscalculations in spending planning) [13].

Will AI save the F-35I? According to the DoD plans [14], in the near future, F-35 pilots will be able to use AI to control a small group of drones flying nearby from the aircraft's cockpit in the air, performing sensing, reconnaissance, and targeting functions. The F-35 maintenance system ALIS includes early applications of AI in which computers perform assessments, go through checklists, organize information, and make some decisions on their own – without human intervention.

Already in 2012 [15], F-35 Program software contained 24 million lines of code. Nowadays the total amount of F-35 Program software is estimated above 80 million lines of code

[16], summing up F-35 onboard software, Joint Simulation Environment, ALIS maintenance tools, and much more – Common Analysis Toolset Data Manager (CATDM) software platform for Industry 4.0 manufacturers. How to get out of this dire situation?

IV. How does GAO estimate the status of AI for Weapon Systems?

For the purpose to overview Artificial Intelligence funds (in accordance with decisions [1] and [2]) the US Government Accounting Office prepared a methodical material [10]. Three waves (types) of AI are identified: Expert knowledge, Machine learning, and Contextual adaptation (Fig. 5).

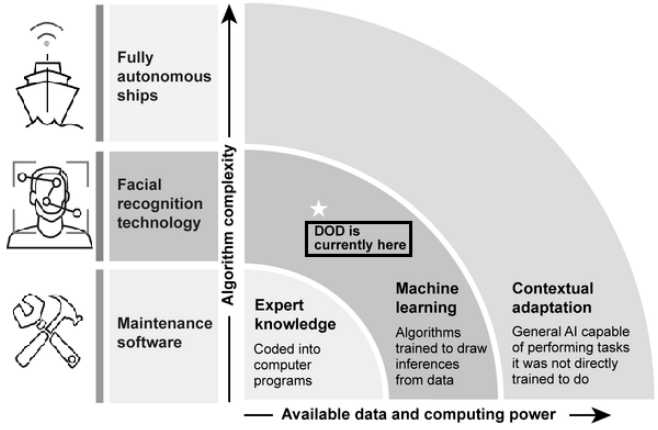

Fig. 5. Types of AI and DOD Examples [10]

Expert knowledge. The first and oldest form of artificial intelligence in which a computer is programmed with detailed rules based on human knowledge or criteria and produces results consistent with its programming.

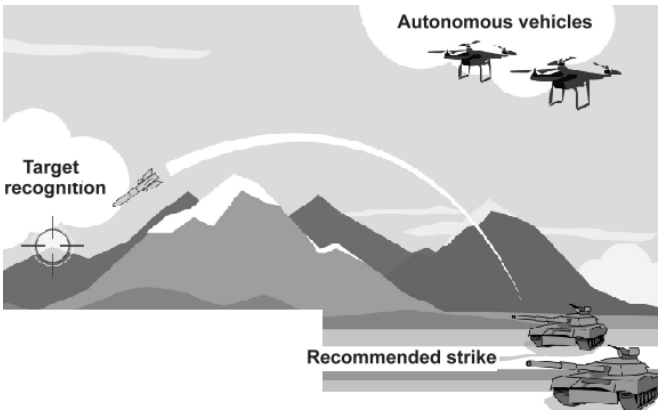

Fig. 6. Example of AI on the Battlefield [10]

An example of such rule-based DOD AI capabilities is aircraft maintenance software that requires users to input information according to predefined data formats and then process that data according to rules programmed by human experts (i.e., maintenance specialists) to diagnose the cause in case of malfunction (Fig. 6).

Let's supplement Fig. 6 with considerations about what the battle process looks like, in which automatic artillery systems participate. Several UAVs are on duty over the battlefield. They carry out artillery reconnaissance, that is, they are engaged in target designation, and this is not new. What is new is that they are able to track the shots that come from the other side. They transmit the coordinates of a flying projectile to a calculator, which instantly determines the trajectory and looks at where the projectile will fall. If he sees that the place where the projectile fell is a threat to one of the artillery installations under the control of the cyber center, then the cyber center simply gives a signal there, and this installation quickly moves away from its place, while the projectile is still flying. The person is excluded from the decision-making process, from the process of controlling the fight. Entirely and completely the battle is conducted by pure automatics - according to clearly defined rules. If the enemy uses a new type of weapon, then without human intervention it is unlikely that defeat will be avoided.

Machine learning. The second and current type of AI, according to GAO experts (see DOD's label, Fig. 5), is the type in which the computer receives basic instructions and training data to learn how to predict specific outcomes, as in Case 2. AI systems drop a challenge to existing Department of Defense assessment strategies and ethical standards for capabilities, which can lead to hesitancy in their use. Figure 7 provides a notional example of AI model complexity and the questions a user may need to be able to answer to trust the AI's decision or recommendation. Ethical standards are a vast, as yet "unplowed" field of international efforts, especially in terms of lethal autonomous weapon systems.

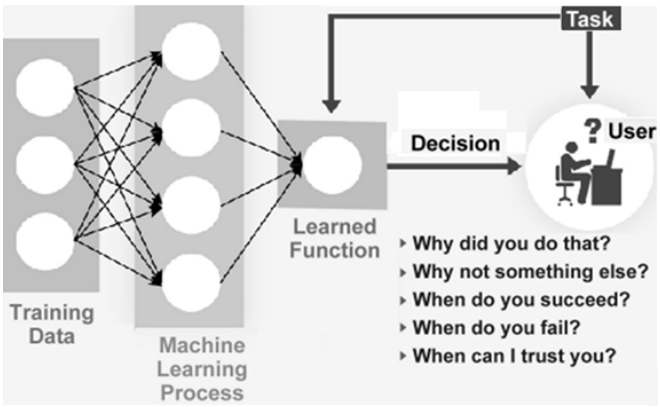

Fig. 7. Example of AI Model Complexity, including ethical issues [10]

Contextual adaptation. A third and as yet unknown potential future type of AI, in which the computer is able to adapt to new situations without the need for retraining, and can also explain to users the reasons for its decisions or predictions. A potential example of a DOD is a fully autonomous ship that uses algorithms to maneuver in situations for which it was not specifically trained (such as inclement weather or contested waters) and is capable of planning, relaying, and carrying out military missions in a manner similar to a way a person would. The majority of such types of AI warfighting capabilities are still in development. These capabilities largely focus on analyzing intelligence, and enhancing weapon system platforms such as aircraft and ships that do not require human operators.

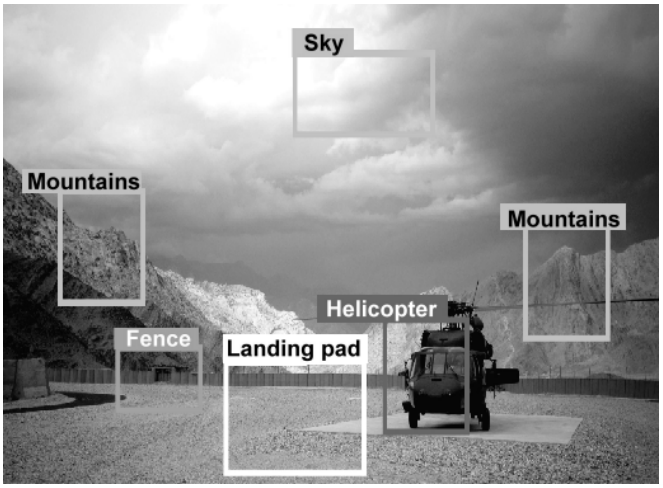

Fig. 8. Example of Labeled Imagery Data [10]

The process of training an AI model is achieved by providing the AI algorithm with large datasets that define the desired outcome, while the AI developer checks if the model produces the desired results. For example, training an AI model to recognize a submarine from a video stream requires a large image dataset of various types of submarines that are identified as submarines. During training, images of submarines will be presented to the system, and personnel involved in the training will confirm when the AI model correctly identifies the submarine and when it does not.

High-performance AI typically requires well-tagged historical data to train the system. Labeled data refers to raw data (images, text files, videos, etc.). Data has been labeled with some identifiers to provide context so that the AI algorithm can learn from it. For example, intelligence, surveillance, and reconnaissance AI capabilities trained using tagged data to identify tanks would require images of various tanks tagged as such. Figure 8 shows another example of such marking. The success of contextual adaptation is not yet clear

V. AI AUDITABILITY – A VERY HARD TASK

Autonomous vehicle. Let's start the talk about the auditability of Artificial Intelligence solutions from the autonomous vehicle revolution: how insurance must adapt [17]. Transportation networks and associated companies are leading a revolutionary shift from individual ownership to new approaches to vehicle mobility and access, including increased use of autonomous vehicles, which Deloitte predicts will account for more than 80% of new vehicle sales in urban areas by 2040. Automotive companies will need to rethink how they manage risk.

Manufacturers, component suppliers, and technology companies involved in building autonomous vehicles and the software that drives them bear a greater risk of liability. The limiting factor on how quickly this position will be adopted by insurers. There is a lack of significant precedents and claims data for autonomous vehicle incidents. What's more, access to the vast amounts of data collected by vehicles that can help determine conditions during collisions is a hard point for insurers.

On the European AI act. While some AI applications are security critical, the use of AI systems creates new challenges regarding aspects such as IT security, safety, robustness, and reliability. Meeting these challenges requires a common framework for auditing AI systems throughout their lifecycle, including assessment strategies, tools, and standards. This is under development, but is only partially ready for practical use at the moment.

In April 2021, the European Commission published a draft regulation on AI (the AI act, AIA) [3]. The goal is to ensure that AI systems in practice fulfill adequate requirements. The AIA takes a risk-based approach that completely bans some AI applications (such as social scoring schemes) and imposes comprehensive requirements on AI systems that are considered high-risk. According to the AIA, high-risk applications include, among others, the use of AI in security-critical functions, as well as in healthcare and justice, and law enforcement.

The AIA [3] proposes ultra-heavy fines of up to €30,000,000 or, if the offender is a company, up to 6% of its total annual worldwide turnover for the previous financial year, whichever is greater. These penalties are applied in cases of:

(a) non-compliance with the prohibition of the artificial intelligence practices (e.g., the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement);

(b) non-compliance of the AI system with the requirements of data governance for high-risk AI systems.

Auditable AI Systems. The research on and application of artificial intelligence (AI) has triggered a wide scientific, economic, social, and political discussion. To implement the AIA directive, it is necessary to develop audit schemes, methods, and tools for all aspects mandated by the AIA across the relevant AI life cycle phases. Such type of research should be done within the next two to three years, when the AIA will start to apply. The status in this area was discussed during the 2nd International Workshop "Towards Auditable AI Systems", on October 26th, 2021, in Berlin, organized by the Federal Office for Information Security Germany, the TÜV-Verband, and the Fraunhofer HHI (see Whitepaper [18]).

AI systems bring up new challenges for auditing as compared to classical software, namely,

(1) the input and state spaces of AI systems for common tasks are enormous, making fool testing infeasible;

(2) their behaviour strongly depends on the data used to train them, and any manipulations of these data can turn to grave consequences;

(3) most AI systems nowadays used to have a complex inner structure that is not able to human interpretation.

Therefore, a very hard task is to find malfunctions and attacks and mitigate them. These challenges are extremely important and to facilitate the AI application in secure, robust, and transparent conditions, especially in security and safety-critical applications, it is necessary to have available strong

technical requirements for auditing AI systems. However, such material is largely missing so far.

Document [18] focuses on connectionist AI systems, which e.g. are used in applications based on image processing. Connectionist AI systems use large data structures. Such structures contain millions of parameters. The currently most widespread examples are deep neural networks (DNNs), which consist of various layers of simple processing elements (neurons) that are highly interconnected (as in Case 2 on leaf diseases above).

The document focuses on those requirements from the AIA that are related to IT security and safety. A two-dimensional "Certification Readiness Matrix" was proposed as a tool to monitor the progress of AI auditability (Fig. 9).

Lifecycle Phase / Aspect	Security	Safety	Performance	Robustness	Interpret-/ Explainability	Tracability	Risk Management
organization	3	2	5	3	4	6	6
use case specific requirements & risks	5	5	5	5	4	4	6
Embodiment & situatedness of AI module	5	5	5	5	6	2	5
planning phase	4	4	5	4	4	6	6
data acquisition and QA phase	4	5	6	6	4	6	6
training phase	5	5	5	5	6	6	6
evaluation phase	5	5	5	5	6	6	6
deployment and scaling phase	4	2	5	3	4	6	6
operational (& maintenance) phase	5	2	5	3	4	6	6

Auditability Scoring: 0 (none) 2.5 5 (average) 7.5 10 (full)

Fig. 9. An example of a certification readiness matrix [18]. The auditability scoring scale shows color and point scales that correspond to a scale from non-existing auditability (red, 0) to full auditability (green, 10).

The "Certification Readiness Matrix" presented here is meant as a conceptual heuristic covering the AI life cycle phases and the embedding of an AI system within organizational processes. There are seven aspects:

- security (passive and active robustness of the AI system against attacks);
- safety (protection against (physical) harm);
- performance (with respect to relevant performance metrics);
- robustness (against natural variations of situations, including those, that were not covered during training);
- interpretability and explainability (the ability of humans to understand the decision process of the AI system);
- the ability to monitor (track) the AI system at all its stages (at all steps of the machine learning pipeline). This includes monitoring design decisions, analyzing initial data (training sets), checking boundary conditions, monitoring system performance, etc.;
- risk management (a minimization of risk probability and/or impact; includes strategic and operational measures).

At the moment, the requirements, audit methods, and audit tools are not sufficiently available, but the development in this area is very dynamic. Supposedly, statistics as an interdisciplinary scientific field can play a significant role both in the theoretical and practical understanding of AI, and for its future development. Statistics can even be considered the main element of AI [19]. How to achieve this? Is it possible to correct the situation by virtue of AIA?

VI. On the Robustness of Machine Learning Software

We turn attention to the ML robustness as a key point relating to AI auditability, on our opinion. The problem of robustness prevents the widespread introduction of machine learning systems in critical areas (avionics, nuclear systems, autonomous driving, etc.).

The definition of robustness (let's call it the Robustness Criterion) is borrowed from mathematics, and, approximately, corresponds to the following form. Given an input x and a model f, we want the model prediction to remain the same for all inputs x' in a neighborhood of x, where the neighborhood is defined by some distance function δ and some maximum distance Δ. That is, the results of the classifier, for example, would not change with a small change in the data. The fundamental basis of robustness research is quite clear. Basically, any model is trained on some subset of data, and then generalized to the entire population of data, which, in general, is unknown at the time of training. Therein lies the issue of robustness. If the data is changed in a special way, then this is called an attack on machine learning systems.

It is around the Robustness Criterion that all research in the field of artificial intelligence is built. How to select minimally different data, which, nevertheless, is classified differently? Since in most cases, we are talking about images, we are talking about changes imperceptible to the human eye that lead to a change in classification. How important is the "invisibility" of changes, if in critical applications (avionics, etc.) we are dealing with automatic systems?

It is assumed that the performance of the model, achieved at the stage of training, is preserved during its practical use. There is a complete parallel with traditional software implementation. During the testing phase, we checked the performance of the system, and we expect this performance to continue during the operational phase. Note that for critical applications, the software is also subject to certification. The meaning of this certification is precisely in comprehensive testing (proof of correct operation). According to the same principle, robustness is perceived. That is, robustness becomes synonymous with performance.

In addition, it is not enough to obtain a formal confirmation (verification) of the model operation [20], since the question of scaling arises. Note that in the classical approach to building mathematical models, complexity was never a virtue, the model had to be as simple as possible. In machine learning models, the number of parameters is already measured in billions and formal methods for checking models (by means of logical statements or solving a system of linear equations) are not acceptable. It turns out that in any case, we will be content with some estimates. And this, in turn, does not correlate well with the fact that software must be certified for critical applications [21], and classical approaches to such certification do not work with non-deterministic systems, which are machine learning applications. Certification of machine learning systems for critical applications is still an open area [22].

Testing of machine learning systems can be solved by means of adversarial attacks [23], which are applicable at all stages of the machine learning process. At this point in time, attacks seem to be ahead of defenses (attacks appear first, and only then appear defenses against them). In fact, it must be recognized that the success of machine learning (and, accordingly, artificial intelligence) today is associated with generative models. With discriminant models today, there is some dead end in terms of critical applications. Results can be obtained, but they cannot be guaranteed in the general case. As the results of many projects on robust machine learning systems [24] show, there are no clear achievements in this area yet.

VII. The gap between machine learning and statistical communities

Different cultures. There is a great difference between machine learning and statistical communities; they have different cultures and different scientific backgrounds [25]. The machine learning community has its roots in engineering, computer science, and especially in artificial intelligence as a kind of neuroscience. The ML community tends towards marketing, publishing, and trying to sell their ideas. This feature reflects in a desire to monetize algorithms in the near term, thus focusing on industry problems rather than scientific problems. The path to monetization in science is often much longer and less assured. A large (if not major) share of machine learning's success must be attributed to its very successful and aggressive marketing efforts.

The statistics community is primarily made up of researchers who received an initial degree in mathematics and graduate training in statistics. Statisticians are not hurrying to publicize their research, and their training tends to differ dramatically from that of ML researchers. Statisticians usually have a strong background in mathematics that includes multivariate calculus, linear algebra, differential equations, and real analysis. They then require years of probability and statistics, namely, of asymptotic theory, statistical sampling theory, hypothesis testing, and experimental design. ML researchers have much less knowledge in many of these areas, but have a stronger background not in just programming, but also in signal processing and computing. Will it be possible to bridge the gap between machine learning and statistical communities?

On neural networks and statistical tools. The paper [26] published in 1996 discusses neural networks and compares them to regression models. A comparison between regression analysis and neural networks in terms of notation and implementation is made to help the reader understand neural networks. This shows that neural networks act as a type of non-parametric regression model, allowing us to model complex functional forms. But these results have not received further work on machine learning. 25 years have passed since the publication of the article [26], and it has been cited more than 550 times, but the gap between machine learning and the statistical communities has only increased. For example, in the encyclopedia [27], statistical cases occur only in passing.

The non-parametric nature of neural networks allows the development of models without any prior knowledge of the distribution of the data set or possible interaction effects between variables, as required by commonly used parametric statistical methods. For example, multiple regression requires that the error in the regression equation be distributed normally. Another statistical technique that is often used for categorization is discriminant analysis, but discriminant analysis requires the predictor variables to be multivariate with a normal distribution. Such type assumptions are removed from AI models. Encyclopedia [27] does not even pose the problem of convergence between machine learning and statistical communities.

Will statistics help? The statistical methods are fundamental for finding structure in data and for obtaining deeper insight into data and having success in data analysis. Ignoring statistical data analytics may lead to avoidable fallacies [28]. This holds, in particular, for the analysis of big data, and the notion of distribution is the key point of statistics. Only the probability distributions allow us to predict error bands. The unfortunate thing is that in the field of AI data distributions are unknown. There remains hope for non-parametric statistics.

Search for non-parametric statistics. Non-parametric (or distribution-free) statistical methods are mathematical approaches for statistical hypothesis testing which, unlike parametric statistics, make no assumptions about the probability distributions of the variables being assessed. The following are the most frequently used tests:

- Anderson-Darling test: whether a sample is drawn from a given distribution;
- Kolmogorov-Smirnov test: whether two samples are drawn from the same distribution;
- Siegel-Tukey test: tests for differences in scale between two groups;
- Sign test: whether pair samples are drawn from distributions with equal medians;
- Spearman's rank correlation coefficient: measures statistical dependence between two variables using a monotonic function;
- Squared ranks test: tests equality of variances in two or more samples.

Will the list of non-parametric statistics be supplemented with values from the ML area? This task is extremely difficult. But without this, it is unlikely that the AIA document will be put into effect.

On statistical paradox in big data. The key mathematical tool for justifying statistical methods is large-sample asymptotics. There are two basic statistical notions: the Law of Large Numbers and Central Limit Theorem. Neither of them could be established without such asymptotics. When statisticians have the explosive growth of data size, they hope to get the large-sample asymptotic results out there. The reality appears to be the opposite. In [29], the author discusses a simple statistical quality control task: "Which one should we trust more, a 5% survey sample or an 80% administrative dataset?"

A deeper study of this task led to the development of the new notion Data Defect Index and to shift from the traditional focus on probabilistic uncertainty in the familiar form

of Standard Error $\propto \sigma/\sqrt{n}$

to the practice of systematic error in non-probabilistic Big Data in an unuseable form

Relative Bias $\propto \rho \sqrt{N}$.

Here "Relative Bias" is the bias in the sample mean relative to a standard error, σ and n are the standard deviation and sample size, and N is the population size. The unfamiliar term ρ is a Data Defect Correlation, defined in [29].

As a big data statistical paradox case, there are estimates obtained from the Cooperative Congressional Election Study (CCES) of the 2016 US presidential election suggest a $\rho \approx -0.005$ for self-reporting to vote for Donald Trump. Because of the Law of Large Populations [29], this seemingly insignificant data defect correlation implies that the simple sample proportion of the self-reported voting preference for Trump from 1% of the US eligible voters, that is, $n \approx 2{,}300{,}000$, has the same mean squared error as the corresponding sample proportion from a genuine simple random sample of size $n \approx 400$, a 99.98% reduction of sample size (and hence our confidence). The CCES data demonstrate the power of LLP: on average, the larger the state's voter populations, the further away the actual Trump vote shares from the usual 95% confidence intervals based on the sample proportions. This should remind us that, regardless of the quality of the data, population inferences from big data are subject to the Big Data paradox: the more data, the more we fool ourselves.

It is possible that the two new concepts of Data Defect Correlation and the Law of Large Populations introduced by the author [17] can serve as the beginning of the search for new measures for Big Data. This relatively recent work has already been cited 170 times. Similar type of researches are collected in [30], and on their basis, it is proposed to develop a new formal methodology. Thus, the Big Data problem can serve as a basis for the fundamental scientific reform.

VIII. Conclusions

This is an insight into the difficulties that arise when looking at applications of artificial intelligence (AI) from traditional statistical data processing. In addition, it is associated with an unprecedented amount of Big Data, including the grandiose amount of software. At first glance, it seems that the reason that causes of Big Data analysis failures is the difference in cultures between machine learning and statistical communities. But the reason is apparently deeper, as the statistical paradox in the Big Data example above shows. At present, it is not clear whether it will be possible to invent parameters that will help meet the requirements of insurance companies for safety- and security-critical Artificial Intelligence (AI) applications. It is possible that the two new concepts of Data Defect correlation and the Law of Large Populations introduced by the author [29] can serve as the starting point of the search for new measures for Big Data. The analysis of AI studies showed that there are currently no indicators that can measure the security, safety, robustness, and trustworthiness of software used in AI applications. We cannot remain silent about the cyber threat situation either, which makes Big Data analysis extremely difficult. The task of providing robustness of machine learning software, especially in safety- and security-critical areas, is currently beyond the competence of individual companies and even governments and is becoming a problem of international cooperation.

References

[1] Exec. Order No. 13,859, 84 Fed. Reg. 3967 (February 14, 2019) https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence Retrieved: Oct, 2022.

[2] U.S. Department of Defense. Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity, February 12, 2019.

[3] European Commission: Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), 2021, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206 Retrieved: Oct, 2022

[4] F.E MORGAN, et al, Military Applications of Artificial Intelligence. RAND, 2020

[5] M. Sloane, E. Moss, R. Chowdhury, "A Silicon Valley love triangle: Hiring algorithms, pseudo-science, and the quest for auditability." *Patterns* 3.2 (2022): 100425

[6] K. P. Ferentinos, "Deep learning models for plant disease detection." *Computers and Electronics in Agriculture* 145 (2018): 311-318

[7] GAO-19-128, Weapon Systems Cybersecurity. DOD Just Beginning to Grapple with Scale of Vulnerabilities, Report to the Committee on Armed Services, U.S. Senate, US GAO, October 2018

[8] D. Grazier, "What Should We Do About a Generation of Weapons Vulnerable to Cyberattacks? An Obvious Solution Being Ignored," January 31, 2019. Web: https://www.pogo.org/analysis/2019/01/what-should-we-do-about-a generation-of-weapons-vulnerable-to-cyberattacks/ Retrieved: Oct, 2022

[9] J. Gill, "JAIC and DDS as offices cease to exist". May 24, 2022. Web: https://breakingdefense.com/2022/05/say-goodbye-to-jaic-and-dds-as-offices-cease-to-exist-as-independent-bodies-june-1/ Retrieved: Oct, 2022

[10] GAO-22-104765. Artificial Intelligence: Status of Developing and Acquiring Capabilities for Weapon Systems. Feb 17, 2022
[11] F-35 program stagnated in 2021. Web: https://www.pogo.org/analysis/2022/03/f-35-program-stagnated-in-2021- but-dod-testing-office-hiding-full-extent-of-problem/ Retrieved: Oct, 2022
[12] GAO-22-105995. F-35 sustainment DOD. Faces Several Uncertainties and Has Not Met Key Objectives, April 28, 2022
[13] D. Grazier, "F-35 Program Stagnated in 2021 but DOD Testing Office Hiding Full Extent of Problem." https://www.pogo.org/analysis/2022/03/f-35-program-stagnated-in-2021-but-dod-testing-office-hiding-full-extent-of-problem Retrieved: Oct, 2022
[14] K. Osborn, "Artificial Intelligence is Going to Make America's F-35 and B-2 Even Stronger", December 14, 2019 https://nationalinterest.org/blog/buzz/artificial-intelligence-going-make-americas-f-35-and-b-2-even-stronger-104967 Retrieved: Oct, 2022
[15] R.N. Charette, "F-35 Program continues to struggle with software." *IEEE Spectrum* 19(2012).
[16] West, T D. and Blackburn M., "Is Digital Thread/Digital Twin Affordable? A Systemic Assessment of the Cost of DoD's Latest Manhattan Project," *Procedia Computer Science* 114 (2017):47–56
[17] Autonomous Vehicle: A New Approach to Insurance. The Autonomous Vehicle Revolution: How Insurance Must Adapt (marshmclennan.com)
[18] Towards Auditable AI Systems. Whitepaper. May 2022. https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Towards_Auditable_AI_Systems_2022.pdf?__blob=publicationFile&v=4 Retrieved: Oct, 2022
[19] S. Friedrich, G. Antes, G., Behr, S. et al, "Is there a role for statistics in artificial intelligence?" *Adv Data Anal Classif* (2021):1-24.
[20] D. Namiot, E. Ilyushin, and I. Chizhov, "On a formal verification of machine learning systems." *International Journal of Open Information Technologies*, 10.5 (2022): 30-34.
[21] N. Berthier, et al. "Tutorials on Testing Neural Networks." arXiv preprint arXiv:2108.01734 (2021).
[22] E. Jenn, et al, "Identifying challenges to the certification of machine learning for safety critical systems", in *European Congress on Embedded Real Time Systems* (ERTS 2020). 2020.
[23] H. Li, D. Namiot, "A Survey of Adversarial Attacks and Defenses for image data on Deep Learning." *International Journal of Open Information Technologies* 10.5 (2022): 9-16.
[24] D. Namiot, E. Ilyushin, and I. Chizhov. "Ongoing academic and industrial projects dedicated to robust machine learning." *International Journal of Open Information Technologies* 9.10 (2021): 35-46. (in Russian)
[25] D. B. Dunson, "Statistics in the big data era: Failures of the machine." *Statistics & Probability Letters* 136 (2018): 4-9
[26] B. Warner, M. Misra, "Understanding neural networks as statistical tools." *The american statistician* 50.4 (1996): 284-293
[27] Walczak, Steven. "Artificial neural networks." *Encyclopedia of Information Science and Technology, Fourth Edition*. IGI global, 2018. 120-131.
[28] C. Weihs, K. Ickstadt, "Data science: the impact of statistics." *Int J Data Sci Anal* 6.3(2018):189–194
[29] Meng, Xiao-Li, "Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election." *The Annals of Applied Statistics* 12.2 (2018): 685-726
[30] B. Devezer, et al, "The case for formal methodology in scientific reform." *Royal Society open science* 8.3 (2021): 200805.
[31] F-35 Joint Strike Fighter (JSF) Program. Congressional Research Service. Updated May 2, 2022. https://crsreports.congress.gov RL30563 Retrieved: Oct, 2022