Intelligence Augmentation and Human-Machine Interface Best Practices for NDT 4.0 Reliability

by John C. Aldrin*

ABSTRACT

NDT 4.0 is a vision for the next generation of nondestructive inspection systems following the expected fourth industrial revolution based on connected cyber-physical systems. While an increasing use of automation and algorithms in nondestructive testing (NDT) is expected over time, NDT inspectors will still play a critical role in ensuring NDT 4.0 reliability. As a counterpoint to recent advances in artificial intelligence algorithms, intelligence augmentation (IA) refers to the effective use of information technology to enhance human intelligence. While attempting to replicate the human mind has encountered many obstacles, IA has a much longer history of practical success. This paper introduces a series of best practices for NDT IA to support NDT 4.0 initiatives. Algorithms clearly have a great potential to help alleviate the burden of "big data" in NDT; however, it is important that inspectors are involved in necessary secondary indication review and the detection of rare event indications not addressed well by typical algorithms. Examples of transitioning algorithms for NDT applications will be presented, emphasizing the successful interfacing of inspector and software for optimal data review and decision making.

KEYWORDS: Industry 4.0, artificial intelligence, intelligence augmentation, human-machine interface, reliability

Materials Evaluation 78 (7): 869–879
https://doi.org/10.32548/2020.me-04133

* Computational Tools, Gurnee, Illinois 60031, USA; aldrin@computationaltools.com

Introduction

Industry 4.0 is a term developed by German industry leaders and researchers to describe how the Internet of Things (IoT), an emerging network of linked cyber-physical devices, will improve engineering, manufacturing, logistics, and life-cycle management processes (Jahanzaib and Jasperneite 2013). The number 4.0 refers to a fourth industrial revolution. Beginning in the 1700s, three major waves of technological changes transformed the industrial landscape and increased productivity: (1) mechanization and water/steam power; (2) mass production (for example, assembly lines) and electricity; and (3) computers and automation. The fourth industrial revolution is expected to be based on connected cyber-physical systems. There is a parallel vision for the next generation of NDT capability referred to as NDT 4.0 (Meyendorf et al. 2017a, 2017b; Link and Riess 2018; Vrana et al. 2018; Singh 2019). A key aspect of NDT 4.0 is leveraging automation in the evaluation of the workpiece and providing characterization of the state of the part for improved life-cycle management (Lindgren 2017; Forsyth et al. 2018).

A diagram of an integrated vision for NDT 4.0 is presented in Figure 1. One key innovation of NDT 4.0 is the integration of advanced control systems and NDT algorithms to support complex inspections, NDT sensor data acquisition, and data analysis tasks. In recent years, major advances have been made in the field of machine learning and artificial intelligence (AI) to perform complex data classification tasks, leveraging training on "big data" sets (LeCun et al. 2015). While this technology is promising, challenges do exist with transitioning machine learning/AI algorithms to NDT applications. Training AI requires very large, well-understood data sets, frequently not available in NDT, and there are major concerns about the reliability and adaptability of such algorithms to completely perform complex NDT data review tasks. One of the primary objectives of this paper is to survey the potential benefits and challenges of emerging algorithms for NDT 4.0 systems. Experience and perspective on the transition of algorithms for NDT applications will also be discussed.

The Inspector and NDT 4.0

A critical component of any NDT tool is the interface with the human that uses it. Figure 1 shows the human-machine interface as a critical link between the NDT inspector/engineer and

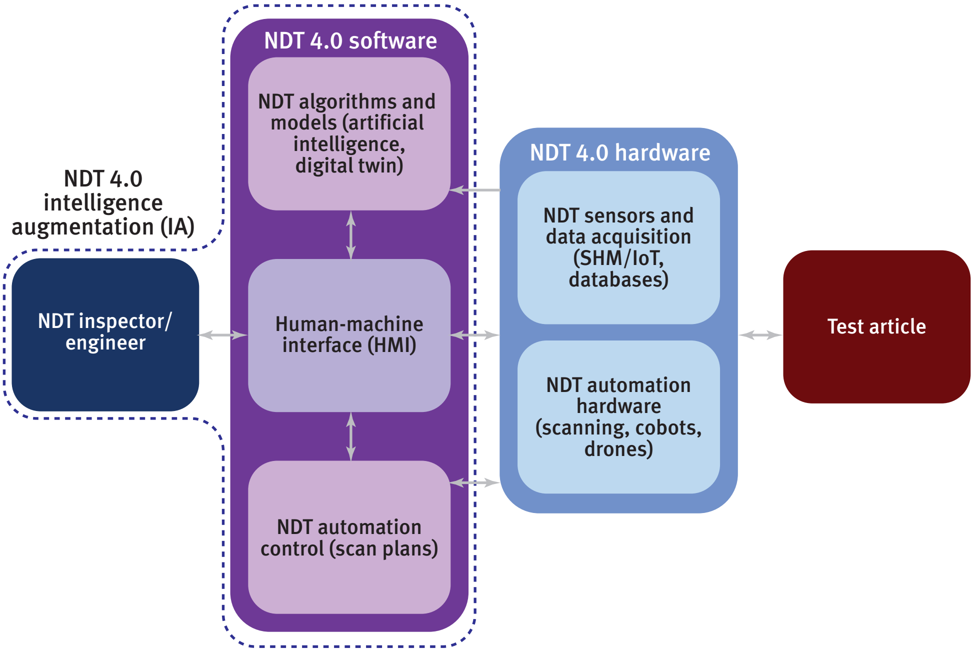


Figure 1. Vision for NDT 4.0: intelligence augmentation (IA) for NDT inspectors and engineers is achieved through a human-machine interface to NDT automation hardware, sensors, and data acquisition algorithms and models.

NDT 4.0 software and hardware. Care must be taken with the implementation of automation to ensure that operators have the necessary awareness and control as needed. In addition, as a counterpoint to recent advances in AI algorithms, intelligence augmentation (IA) is introduced as the effective use of information technology to enhance human intelligence. From this perspective, the inspector is an integrated part of NDT 4.0 systems and performs necessary tasks in collaboration with automated NDT systems and data analysis algorithms. This paper will present a series of best practices for the interface between NDT hardware, software, and algorithms and human inspectors and engineers to ensure NDT 4.0 reliability.

Algorithms and AI in NDT

NDT algorithms that perform indication detection and characterization can be organized into three classes: (1) algorithms based on NDT expert knowledge and procedures (heuristic algorithms); (2) model-based inversion; and (3) algorithms incorporating statistical classifiers and/or machine learning. The most basic algorithm is one based on human experience. The term *heuristic algorithm* is useful to describe a class of algorithms based on learning through discovery and incorporating rules of thumb, common sense, and practical knowledge. This first class of algorithms essentially encodes all key evaluation steps and criteria used by operators as part of a procedure into the algorithm. The second class of algorithms is a model-based inversion that uses a "first principles" physics-based model with an iterative scheme to solve characterization problems. This approach requires accurate forward models and iteratively compares the simulated and measurement data, adjusting the model parameters until agreement is reached. The third class of algorithms covers statistical classifiers and machine learning, which are built through the fitting of a model function using measurement "training" data with known states. Statistical representation of data classes can be accomplished using either frequentist procedures or Bayesian classification. Machine learning and AI are general terms for the process by which computer programs can learn. Early work on machine learning built upon emulating neurons through functions as artificial neural networks using layered algorithms and a training process that mimics a network of neurons (Fukushima and Miyake 1982). In recent years, impressive advances have been made in the field of machine learning, primarily through significant developments in deep learning neural network (DLNN) algorithms (Hinton et al. 2006; LeCun et al. 2015; Lewis-Kraus 2016). Large sets of high-quality, well-characterized data have been critical for the successful training of DLNNs. As well, software tools have been developed for training neural networks that better leverage advances in high-performance computing. A recent overview on algorithms for NDT classification is summarized in a previous paper (Aldrin and Lindgren 2018).

Benefits of Algorithms/AI in NDT

There are a number of advantages associated with incorporating algorithms as part of an NDT technique. First, algorithms are typically very good at performing laborious and repetitive tasks. For most parts under test, either in manufacturing or in service, the presence of critical NDT indications is a fairly rare event. Therefore, the data review process can often be a tedious task for most operators, who can expect mostly good parts. Second, given the amount and complexity of some data review tasks performed for some inspections, such tasks can be a challenge, especially for inexperienced inspectors or inspections that are rarely performed. This trend appears to be growing with the increasing quantity of data acquired with automated scanning and array sensing systems. Third, in many instances, algorithms can perform the data review task faster than manual review, providing potential savings in maintenance time and costs. Fourth, algorithms are typically not biased by expectation, such as the frequency of indications in past inspections. With a reduction in errors, the overall risk of maintaining a component can be improved. Fifth, algorithms can be designed in such a way to support the operator as a "digital assistant." Algorithms could potentially help alleviate the burden of "mostly good data" and allow operators to focus on key data review tasks. As well, algorithms can be used to reduce the size and dimensionality of NDT data and present the operator with a reduced feature set for manual classification. Lastly, there are challenges with the aging workforce and transitioning expert knowledge to the next generation. Algorithms, if designed properly, can be repositories for expert knowledge in an NDT organization.

Challenges of Algorithms/AI in NDT

While the application of NDT algorithms shows great promise, there are a number of potential disadvantages with applying algorithm-based solutions to NDT inspection problems. First, the development and validation of reliable algorithms for NDT can be expensive. Training DLNNs requires very large, well-understood data sets, which are frequently not readily available for NDT applications. While the NDT community often possesses a large amount of data, the material state behind the data is often not perfectly known. Acquiring data from parts with well-characterized damage states, such as cracks, corrosion, or impact damage, requires either high-resolution NDT techniques for fingerprinting, or destructive characterization for full verification. The design, training, and validation of algorithms also require unique software development skills and many hours of engineering labor to successfully implement.

Second, algorithms also can perform poorly for scenarios that they are not trained to interpret. There have been concerns for decades about the reliability and adaptability of machine learning algorithms to completely perform complex NDT data review tasks. In NDT, many promising demonstrations have been performed by the NDT research community, but frequent issues concerning overtraining and robustness to variability for practical NDT measurements outside of the laboratory have been noted (Aldrin and Lindgren 2018). Prior successful NDT applications of neural networks have been dependent on taking great care to reduce the dimensionality of the data and provide reliable features as inputs for classification. As well, designing algorithms to address truly rare events—so-called black swans—is extremely difficult (Taleb 2007).

Third, while human factors are frequently cited as being sources for error in NDT applications, humans are inherently more flexible in handling unexpected scenarios and can be better at making such judgement calls. Human inspectors also have certain characteristics like common sense and moral values, which can be beneficial in choosing the most reasonable and safest option. In many cases, humans can detect when an algorithm is making an extremely poor classification due to inadequate training and correct those errors.

Fourth, for many machine learning algorithms like DLNNs, it can be difficult to ascertain exactly why certain poor calls are made. These algorithms are often referred to as "black boxes," because the complex web of mathematical operations optimized for complex data interpretation problems does not generally lend itself to reverse engineering. Approaches are being developed to sample the parameters space to ascertain the likely source for decisions (Olden and Jackson 2002), but the field of "explainable AI" (XAI) is still in its infancy (Stapleton 2017).

Lastly, with the greater reliance on algorithms, there is a concern about the degradation of inspector skills over time. As well, there is a potential for certain organizations to view automated systems and algorithms as a means of reducing the number of inspectors. However, many of these disadvantages can be mitigated through the proper design of human-machine interfaces.

NDT Intelligence Augmentation

With recent progress and hype on the coming wave of AI, some perspective is needed to understand how exactly these algorithms will be used by humans. While the original vision for AI was to mimic human intelligence, in practice AI has been successful only for very focused tasks. While today certain algorithms can perform better than humans for certain predefined and optimized tasks, we have not achieved the early goal of independent AI. Humans not only have the capability to perform millions of different tasks, many in parallel run by the unconscious mind, but they also have the wherewithal to determine when it is appropriate to switch between tasks and allow the conscious mind to have awareness as needed. The real value of AI today is using it as a specific tool (Aldrin et al. 2019).

As a counterpoint to AI, IA refers to the effective use of information technology to enhance human intelligence (Skagestad 1993; Rastogi 2017). This idea was proposed in the 1950s and 1960s by early cybernetics and computer pioneers. IA uses technology to essentially "support" a human in performing specific tasks. Relative to AI, IA has a long history of success. For example, consider the history of information technology, from the birth of writing and slide rules to smartphones and the internet. All of these forms of technology have essentially been developed to extend the information storage and processing capabilities of the human mind. Fundamentally, progress on AI algorithms should be viewed as an evolution of tools to better support human performance.

While most of the attention in recent years has been on the performance of AI over humans in games such as chess and Go (Lewis-Kraus 2016), there are a number of applications that have been cited where humans plus algorithms can exceed the performance of computer algorithms alone. "Centaurs" (Scharre 2016; Case 2018) and "cyborgs" (Tharp 2017) are terms used to refer to such human-plus-machine collaborations. One example that is frequently cited is chess. A team of amateur chess players paired with three chess programs convincingly defeated a series of teams made up of chess grandmasters and some of the world's best chess programs (Cowen 2013; Tharp 2017). While this case study is slightly dated and may not hold up to the success of AlphZero (Gerbert 2018), fundamentally, all of these algorithms at some stage in their design for operational tasks have incorporated human input. This collaboration between humans and algorithms leveraging high-performance computing has the potential to solve an array of greater problems than mere games of strategy. For example, for many decades the practice of engineering has consisted of humans leveraging their intellect with the support of computational tools to solve technical problems. Humans are still critical in asking the right questions and providing the appropriate focus, complementing the brute force computational power with creativity in selecting the most promising problem space to investigate (Wilson and Daugherty 2018). Humans also have a natural flexibility, versatility, and intuition that AI systems have yet to achieve. These uniquely human qualities are still quite impressive, especially considering the relatively low power consumption of the human mind.

From the perspective of NDT applications incorporating algorithms, IA has the potential to address most of the disadvantages of the AI-based algorithms cited previously. For example, many of the most promising DLNN applications today—from speech recognition to text translation and image classification—are still far from perfect. However, that does not mean that these tools are not useful. In practice, humans can frequently detect errors made by AI and can quickly work around poor results. Humans often develop an understanding where such algorithms can be most appropriately applied and where they should be avoided. By leveraging the algorithms where they are most useful, it becomes less critical for the algorithm to be able to handle all scenarios, especially very rare events. Lastly, by operators working in conjunction with algorithms, there is no need to pursue eliminating the human entirely. In general, the most cost-effective and reliable solution will mostly likely be some hybrid, human-plus-machine based approach.

Human-Machine Interfaces

Typical human interfaces with computer systems in NDT have included monitors, keyboards, mouses, and possibly joystick interactions. While these classic PC interfaces are still efficient for many tasks, there are also a number of emerging devices and tools that connect humans with automation. For example, industrial touchscreen tablets, augmented reality glasses, wearable devices (such as smartwatches), voice-recognition systems, and position tracking devices (such as Microsoft Kinect) all have the potential to provide more natural human-machine interfaces to support emerging NDT 4.0 systems. Several promising applications of augmented reality for aircraft maintenance applications have demonstrated feasibility in recent years (Avatar Partners 2017; Jordon 2018). Unique visualization support tools have also been developed for automatically aligning and visualizing data to 3D models, which enables detailed analysis to detect trends at specific locations on the model, indicating potential process problems (Sharp et al. 2009).

Challenges for Implementation

While this is an exciting time for new human-machine interface tools, there is a critical need to carefully optimize the fine interactions between humans and computer algorithms in NDT. Some work has studied the human-machine problem for different NDT applications (Dudenhoeffer et al. 2007; Bertović 2016a, 2016b). For example, Bertović performed a detailed survey of prior work on human factors when interfacing with automation in NDT (Bertović 2016a). While extensive human-automation interaction has clear benefits, research suggests that increased automation has a number of challenges, costs (a paradox frequently dubbed as "automation ironies" [Bainbridge 1987]), or "automation surprises" (Sarter et al. 1997). In this work, a failure modes and effects analysis (FMEA) was conducted to identify potential risks, and a number of preventive measures were proposed. Subsequent studies were used to verify the benefit of the preventive measures, highlighting mixed levels of success (Bertović 2016a).

Additional guidance on the challenge of human-machine interfaces can be gained from the experiences of other communities that also require very high levels of reliability. For example, in aviation, the use of autopilot systems and the handoff between human control and autopilot is a pertinent case study for NDT. In recent years, the accident rate for major aircraft has been reduced to one major accident per

2.56 million flights (Oliver et al. 2017). While overall air safety has been improving, incidents of the loss of control have not. Loss of control occurs when pilots fail to recognize and correct a potentially dangerous situation, causing an aircraft to enter an unstable condition. One example of the possible catastrophic consequences of automation is the tragic crash of flight AF447 (Oliver et al. 2017). Another very recent example concerns the Lion Air and Ethiopian Airlines crashes of Boeing 737 Max aircraft (Rice and Winter 2019). Such incidents are typically triggered by unexpected, unusual events, often comprising multiple conditions that rarely occur together, that fall outside of the normal repertoire of the pilot's experience. In the case of the Boeing 737 Max, the handoff between the human and autopilot systems appears to have not been designed well, and many pilots were not instructed on its operation (Rice and Winter 2019). This paradox of almost totally safe systems, where the same technology that allows the systems to be efficient and largely error-free, can also create systemic vulnerabilities that result in occasional catastrophes. Lessons learned from these cases include: (1) avoiding the cycle of implementing more automation to correct for poor human performance with existing autopilot systems; (2) encouraging more hand-flying to prevent the erosion of basic piloting skills; (3) improving the management of handovers from machines to humans; (4) increasing pilot training for rare events; and (5) supplementing training using simulation of various rare event scenarios. It is critical to avoid the natural tendency to blame the human in these situations when the human-machine interface and/or algorithm design is poor (Hao 2019). Alternatively, it is important to find ways to make such systems robust and ideally "anti-fragile" to randomness and disorder in the environment (Taleb 2012). The human operator must have some level of "skin in the game" (Taleb 2012) and not become reliant on automation over time. Designing human-machine interfaces and providing the necessary training to achieve this balance is far from trivial.

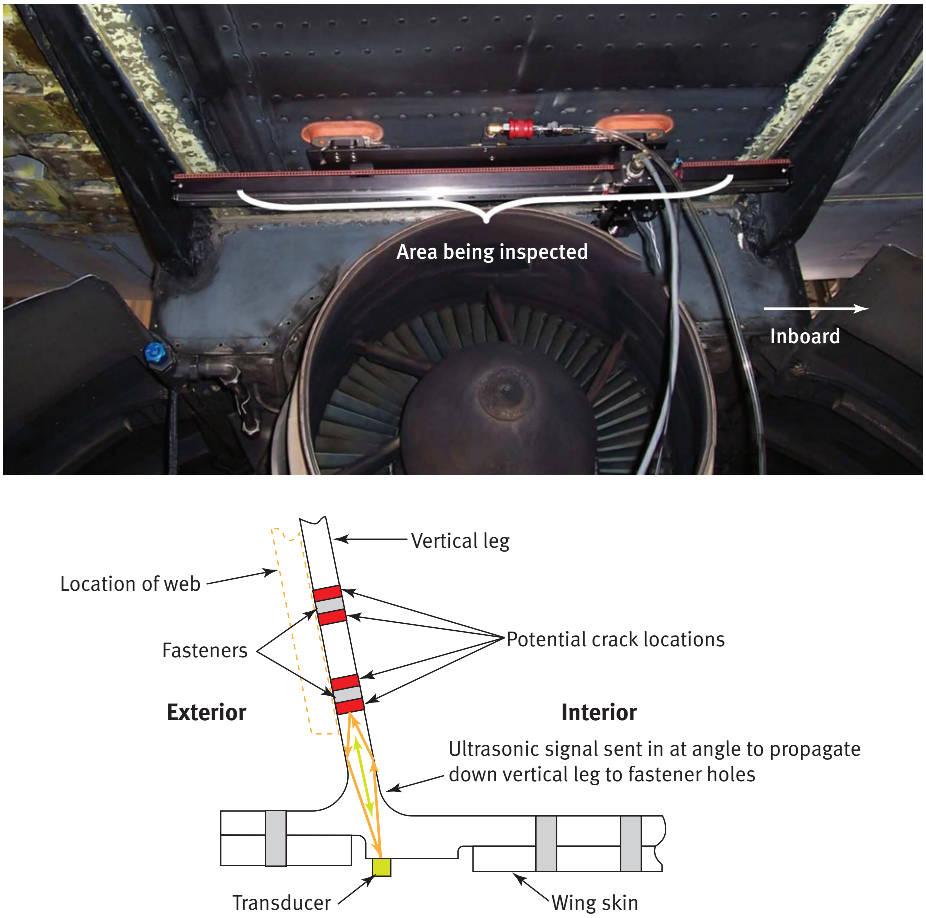
Best Practices for Design of IA and Human-Machine Interfaces

Building on this prior work and experience, a series of best practices for IA in NDT 4.0 is proposed, highlighting how the operator should best interface with NDT data and algorithms. Algorithms clearly have a great potential to help alleviate the burden of big data in NDT; however, it is important that operators are appropriately involved in secondary indication review and the detection of rare event conditions. The following best practices are proposed:

- Provide inspectors with a natural user interface for NDT workflow management. Usability of human-machine interfaces is a critical aspect of workflow management for NDT techniques, from setup, standardization, data acquisition, and indication review. Ideally, inspectors need a way to report results and efficiently provide feedback on indications. Frequently, there are means in NDT software systems to annotate indication results; however, making this metadata readily available to external systems is one of the challenges for NDT 4.0 going forward. Such information will be very useful for refining NDT algorithms and improving life-cycle management.
- Implement data analysis algorithms to address frequent NDT calls and complex data interpretation. It is important to address the low-hanging fruit on implementing algorithms for NDT applications and to help alleviate the burden for inspectors of reviewing "mostly good" data. As well, some complex interpretation problems (especially in ultrasonic NDT) can benefit from algorithms and data guides. The design of these algorithms requires a focus on the base capability for making NDT indication calls to provide value and help ensure reliability. The algorithm design process should consider the necessary engineering development time, cost for acquiring necessary data, and the approach with the highest likelihood of success. There will be a payoff for some applications, but not all applications may benefit from automation. However, as NDT 4.0 systems mature, development costs for each new application should be reduced.
- Ensure inspectors provide a secondary review of indications and review data for rare events. While there is often an initial desire to have NDT algorithms make all indication calls and present simple (good or bad) calls, based on prior experience, additional information is always requested by engineering and management to understand the details on why an indication call was made. Inspectors need a natural user interface to review each call with supporting data and provide feedback on the call details in light of the technical requirements. As well, because no algorithm will be perfect, inspectors need to have a straightforward means to review NDT data quickly. This entails identifying rare indications and determining when the acquisition of the NDT data is out of specification.
- Develop an integrated NDT "simulator" to provide operator training and support complex indication review. There is a potential to leverage the same software interface for training purposes, by having the operators periodically train and test their skills with various conditions in NDT data. Specific rare events can be stored and introduced periodically as part of the regular re-training process. Thus, the interface could be used similar to how flight simulators are used for pilots to verify their performance under standard conditions and rare events. As well, integrated models within the user interface can also provide a means for the verification of indications and support sizing by the inspectors.
- Implement open architecture for NDT data and reporting. Promising software tools exist to support NDT practitioners with data archiving, visualization, and special queries (Sharp et al. 2009), and continued improvements with usability and functionality are expected in the future. Ideally, to share data between NDT 4.0 components, leveraging open data

standards (such as DICONDE and HDF5) and incorporating flexible software architectures, will greatly accelerate the evolution of these systems (Meier et al. 2017; Vrana et al. 2018).

- Reliability must be demonstrated for NDT 4.0 systems. The capability of inspection procedures incorporating NDT 4.0 systems that depend on the performance of both algorithms and the NDT inspector must be evaluated jointly. Probability of detection (POD) evaluation procedures, such as *MIL-HDBK-1823A* (US DOD 2009), are designed to validate the reliability of NDT techniques, regardless of how the indication call is made.
- Software and algorithms can also support NDT reliability as process controls. Simply demonstrating POD capability does not ensure reliability of the technique (Rummel 2010). FMEA should be performed for all NDT techniques incorporating automation to understand the potential sources for poor reliability (Bertović 2016a). In practice, NDT reliability depends on a reproduceable calibration procedure and a repeatable inspection process (Rummel 2010). Process controls and algorithms can thus be used to ensure all calibration indications are verified and to track key metrics that show the NDT process is repeatable over time and under control. As an example, recent work on model-based inverse algorithms with eddy current inspections has shown the potential to reduce error due to variability in probes through calibration process controls (Aldrin et al. 2017). NDT 4.0 systems are also expected to improve the safety of inspections in dangerous environments. By collecting environmental conditions (using environmental sensors and/or weather monitoring) and test system state data from the site, one can ensure the reliability of the inspection task and reduce the level of risk for all involved.
- Build trust over time and consider the cost-benefit for future algorithms and user interface enhancements. Managing costs and mitigating risk drive most decisions for NDT today. For organizations that depend on NDT, there are likely certain applications that will provide the greatest payoff in terms of cost and quality for their customers, transitioning from conventional NDT



(a)

(b)

Figure 2. Inspection of beam cap holes in C-130 aircraft: (a) photo of area being inspected (looking forward); and (b) diagram of inspection problem (from Lindgren et al. 2005).

to NDT 4.0. The transition of algorithms should initially be a phased approach, to both validate the algorithm's performance and build an understanding of where algorithms are reliable and where limitations exist. By tracking called indications over time, it becomes feasible to refine algorithms as necessary. Building that experience internally and achieving an initial payoff will lead to a broader transition of these best practices across an organization and greater shareholder value. Organizational change management must ease this transition through the proper training of inspectors and also management of expectations.

Applications

Several case studies are presented in the following sections that highlight these best practices of leveraging algorithms in NDT applications and addressing human-machine interfaces. These early examples can be considered in the context of a minimal viable product, providing a product with just enough features to satisfy early requirements and provide feedback for future product development. These examples provide key insight on both the promise for NDT 4.0 applications as well as opportunities for future improvement.

Early Example Where AI Vision Becomes IA in Practice

Following the success of the C-141 weep hole inspection program (Aldrin et al. 2001), the development of automated data analysis algorithms was investigated for the inspection of beam cap holes in US Air Force (USAF) C-130 aircraft (Figure 2a) (Lindgren et al. 2005). Here, the fastener sites of interest were in locations of limited accessibility from the external surface and contain fasteners with sealant (Figure 2b). Due to limitations with the NDT capability at the time, there was a need to develop improved ultrasonic techniques to detect fatigue cracks at these locations. A key challenge was the ability to discern multiple signals originating from a possible crack and a geometric feature in a part that was either closely spaced or superimposed in time. The C-130 beam cap holes provided a special challenge given the skewed riser, installed fasteners, and limited transducer accessibility of the B-scan inspection (Figure 2b). This inspection problem frequently produced reflections from the fastener hole (referred to as reradiated insert signals) occurring at similar times of flight (TOF) as near and far crack signals. To address this challenge, a novel feature extraction methodology was developed to detect the relative shift of signals in time for adjacent transducer locations due to differing echo dynamics from cracks and part geometries (Aldrin et al. 2006). This technique was the first ultrasonic NDT method using assisted data analysis methods, validated through a POD study, to inspect for fatigue cracks on USAF structures (Lindgren et al. 2005).

A view of the operator's user interface, dating back to 16 years now, is presented in Figure 3. The original vision for the approach was to have the automated data analysis (ADA) algorithms make all of the indication calls. The team referred

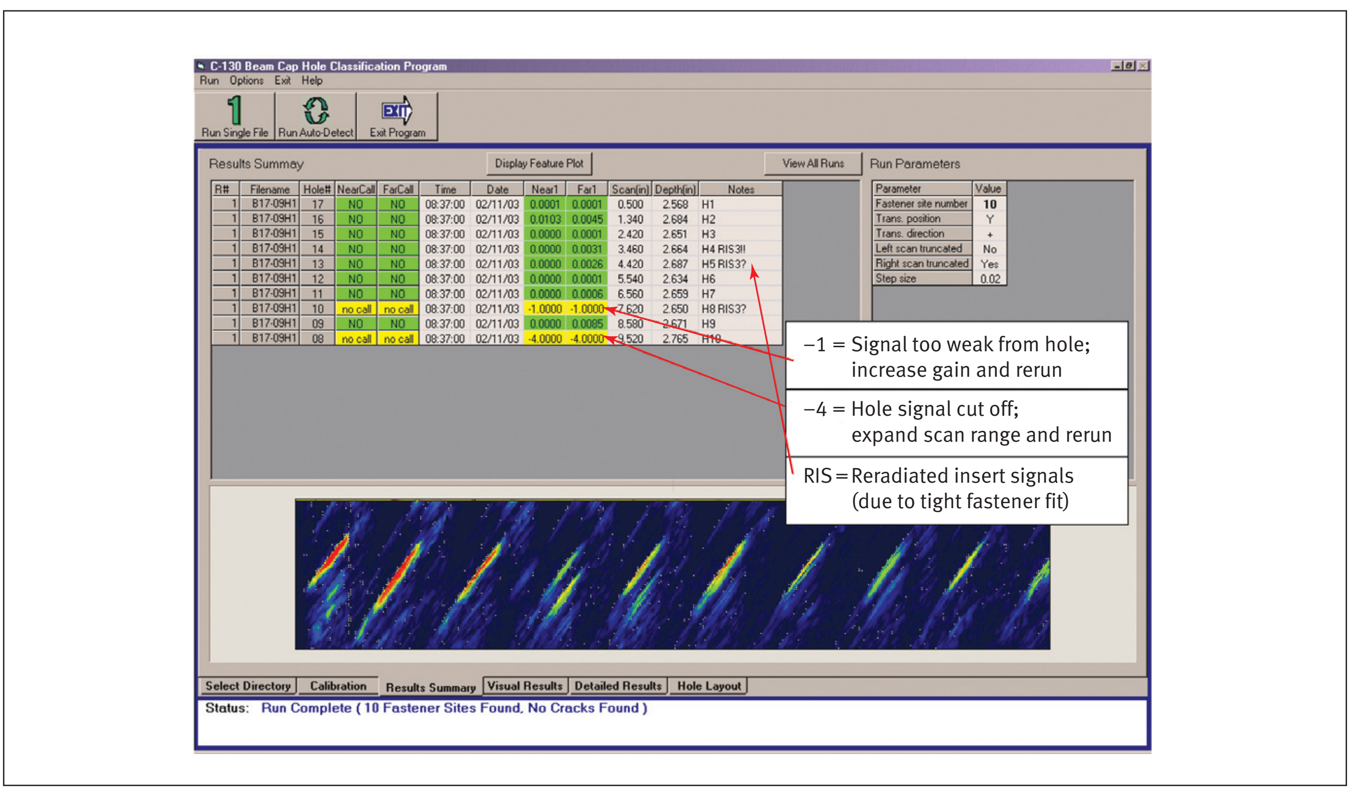


Figure 3. Graphical user interface for automated data analysis (ADA) software incorporating neural network classifiers.

to this early interface as simply "red light/green light." However, during transition, when a call was made by the algorithm, the question of "Why was the call made?" would immediately follow. Enhancements were made to the software to provide more specific feedback on called indications and highlight when data was not adequate for making indication calls. A couple examples of hole indications being too weak or cut off to make proper calls are shown in Figure 3. As well, certain severe structure-plus-fastener conditions were found to produce false calls on rare occasions. To manage these false calls by the algorithm, the results and raw data required a secondary review by inspectors. Inspectors were trained on what to look for in the B-scan to manage this limitation with the algorithm. Although this technique was the first AI/neural network–based approach used to inspect a portion of the USAF C-130 fleet, this case study is actually a very good example of IA in practice.

Lessons Learned on Improving the Human-Machine Interface

Ultrasonic testing is one of the most effective methods to detect critical defect types and ensure the reliability of aerospace polymer matrix composite structures. Most inspection applications of composites are based on pulse-echo ultrasonic testing and manual C-scan data interpretation. Using amplitude and TOF C-scan data, delaminations, disbonds, porosity, and foreign materials can be detected and located in depth. However, the ultrasonic inspection of large composite structures requires a significant work force and production time. To address this inspection burden, ADA software tools were developed and implemented (Aldrin et al. 2016a). The ADA minimizes the inspector's burden on performing mundane tasks and allocates their time to analyze data of primary interest. When the algorithm either detects a feature in the data that is unexpected or that is found to be representative of a defect, then the indication is flagged for further analysis by the inspector.

A software interface for the ADA toolkit is shown in Figure 4. The main view provides a summary of the found indications in the analyzed data, a visual presentation of an indication map, and quantitative metrics assisting the operator in understanding why each call was made. An example of ADA processing results is reported in the interface display shown in Figure 4. Options are provided to enter feedback into the "review" column to indicate if certain calls are incorrect. This example specimen contains artificial defects that have been added at varying locations and ply depth, including above and below the adhesive layer. Indications are listed in the spreadsheet display in the upper left, and corresponding numbers are presented identifying the indications in the C-scan image display on the right. For these ADA evaluations of the two different scan orientations, the three triangular inserts in the bond region were all correctly called. The left-most triangle is in front of the bond and the right two triangles are behind the bond. Indications for the six inserted

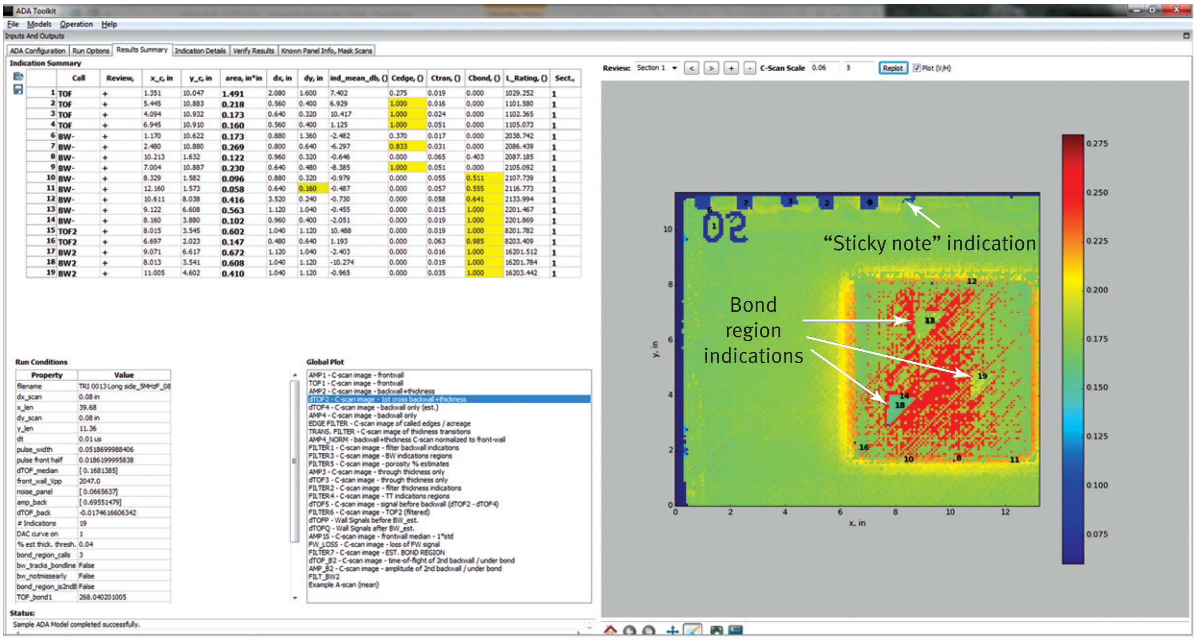


Figure 4. Example ADA toolkit interface with results for a test panel with embedded artificial defects, scanned from tool side, with time of flight C-scan view (adapted from Aldrin et al. 2016a).

materials at the radii are also observed in the TOF map in Figure 4. As well, there are options to add uncalled indications as missed calls into the ADA report with comments. Lastly, features are also provided to support verification of calibration scans, detecting the file and matching the indication calls with expected calibration results. During this development program, feedback from the team and end users was critical in delivering the necessary capability.

Ensuring NDT 4.0 Reliability

A helpful model to represent the reliability of an inspection system has been introduced by Müller and others (Müller et al. 2000, 2014). Total reliability of an NDT system is defined by the intrinsic capability of the system, providing an upper bound for the technique under ideal conditions, with contributions from application parameters, the effect of human factors, and organization context that can degrade the performance. While NDT 4.0 systems are expected to enhance POD performance through improved human factors (supporting ease of use) and repeatability in making complex calls with varying application parameters, NDT 4.0 capability must be evaluated. NDT techniques, whether incorporating AI algorithms, manual inspector data review, or a mixed IA-based approach, require validation through a POD evaluation. Comprehensive POD evaluation procedures (US DOD 2009; ASTM 2015) have been developed to validate the reliability of NDT techniques, regardless of how the indication call is made. In prior work, a POD study was conducted to evaluate the capability of an ADA algorithm to detect cracks around holes in vertical riser aircraft structures (Aldrin et al. 2001). In the study, an ADA approach incorporating neural networks was compared with manual data review by inspectors. Results demonstrated that the automated neural network approach was significantly improved in both detectability, false call rate, and inspection time relative to manual data interpretation (Aldrin et al. 2001). Other recent studies have also addressed the role of POD evaluation when human factors are involved (Bato et al. 2017).

The greatest challenge with validation of NDT algorithms is ensuring that the algorithm is not overtrained but can handle the variability of practical NDT measurements outside of the laboratory. Testing algorithms with independent samples with respect to training data is critical. Model-assisted approaches for training (Fuchs et al. 2019) and validation (Aldrin et al. 2016b) will help provide a diversity of conditions beyond what is practical and cost-effective with experimental data only. Because of these challenges, properly validating NDT techniques using IA is expected to be far easier than a purely AI-based technique. For the example of validating self-driving car technology, simply augmenting the driver experience with collaborative safety systems is much more straightforward to validate than fully validating an AI only–based self-driving car technology. Recent accidents during the testing of self-driving cars indicate the care that is needed to properly and safely validate such fully automated systems when lives are at stake.

Lastly, at this early stage in the application of AI and IA, there are currently no certification requirements for people who design and/or train such algorithms. However, as the field matures, such best practices should be shared throughout the community and included in accredited training programs. Over time, the potential value of implementing certification programs should be considered, possibly under the umbrella of NDT engineering.

Conclusions and Recommendations

While an increasing use of automation and algorithms in NDT is expected over time, NDT inspectors will play a critical role in ensuring NDT 4.0 reliability. As a counterpoint to AI, IA was presented as an effective use of information technology to enhance human intelligence. Based on prior experience, this paper introduces a series of best practices for IA in NDT, highlighting how the operator should ideally interface with NDT data and algorithms. Algorithms clearly have a great potential to help alleviate the burden of big data in NDT; however, it is important that operators are involved in both secondary indication review and the detection of rare event indications not addressed well by typical algorithms. In addition, IA provides more flexibility with the application of AI. When applications are not a perfect fit for existing AI algorithms, a human user can adapt and leverage the benefits of AI appropriately.

Future work should continue to address the validation of NDT techniques that leverage both humans and algorithms for data review and investigate appropriate process controls and software design to ensure optimal performance. Currently, AI algorithms are being developed primarily by engineers to perform very specific tasks, but there may come a time soon when AI tools are more adaptive and offer collaborative training. It is important for adaptive AI algorithms to maintain a core competency while also providing flexibility and learning capability. Care must be taken to avoid having an algorithm "evolve" to a poorer level of practice, due to bad data, inadequate guidance, or deliberate sabotage. Like computer viruses today, proper design practices and FMEA are needed to ensure such algorithms are robust to varying conditions. It is important to design these systems to periodically do self-checks on standard data sets, similar to how inspectors must verify NDT systems/transducers using standardization procedures or having inspectors perform NDT examinations periodically.

Lastly, NDT 4.0–connected initiatives such as digital threads and digital twins are examples of how material systems can be better managed in the future (Kobryn et al. 2017; Lindgren 2017). The digital thread provides a means to track all digital information regarding the manufacturing and sustainment of a component and system, including the material state and any variance from original design parameters. The digital twin concept provides a digital equivalent of a system and exercises the digital twin model through various use scenarios to evaluate individual performance and forecast possible emerging maintenance issues. NDT 4.0 systems are critical to achieving these digital thread and digital twin concepts, enabling an evolution in knowledge management for end users.

ACKNOWLEDGMENTS

The author would like to acknowledge support for portions of this work from the Air Force Research Laboratory (AFRL) under a SBIR Phase II Contract FA8650-13-C-5180 and under Research Initiatives for Materials State Sensing (RIMSS) II. I would like to thank Eric Lindgren and John Welter of the AFRL and David Forsyth of TRI/Austin for their helpful technical discussions.

REFERENCES

Aldrin, J.C., J.D. Achenbach, G. Andrew, C. P'an, B. Grills, R.T. Mullis, F.W. Spencer, and M. Golis, 2001, "Case Study for the Implementation of an Automated Ultrasonic Technique to Detect Fatigue Cracks in Aircraft Weep Holes," *Materials Evaluation*, Vol. 59, No. 11, pp. 1313–1319.

Aldrin, J.C., C.V. Kropas-Hughes, J. Knopp, J. Mandeville, D. Judd, and E. Lindgren, 2006, "Advanced Echo-Dynamic Measures for the Characterisation of Multiple Ultrasonic Signals in Aircraft Structures," *Insight*, Vol. 48, No. 3, pp. 144–148.

Aldrin, J.C., D.S. Forsyth, and J.T. Welter, 2016a, "Design and Demonstration of Automated Data Analysis Algorithms for Ultrasonic Inspection of Complex Composite Panels with Bonds," *42nd Annual Review of Progress in Quantitative Nondestructive Evaluation, AIP Conference Proceedings*, Vol. 1706, No. 1, p. 020006.

Aldrin, J.C., C. Annis, H.A. Sabbagh, and E.A. Lindgren, 2016b, "Best Practices for Evaluating the Capability of Nondestructive Evaluation (NDE) and Structural Health Monitoring (SHM) Techniques for Damage Characterization," *42nd Annual Review of Progress in QNDE, Incorporating the 6th European-American Workshop on Reliability of NDE, AIP Conference Proceedings*, Vol. 1706, p. 200002.

Aldrin, J.C., E.K. Oneida, E.B. Shell, H.A. Sabbagh, E. Sabbagh, R.K. Murphy, S. Mazdiyasni, E.A. Lindgren, and R.D. Mooers, 2017, "Model-Based Probe State Estimation and Crack Inverse Methods Addressing Eddy Current Probe Variability," *43rd Annual Review of Progress in QNDE, AIP Conference Proceedings*, Vol. 1806, No. 1, p. 110013.

Aldrin, J.C., and E.A. Lindgren, 2018, "The Need and Approach for Characterization - US Air Force Perspectives on Materials State Awareness," *44th Annual Review of Progress in QNDE, AIP Conference Proceedings*, Vol. 1949, No. 1, p. 020004.

Aldrin, J.C., E.A. Lindgren, and D. Forsyth, 2019, "Intelligence Augmentation in Nondestructive Evaluation," *45th Annual Review of Progress in QNDE, AIP Conference Proceedings*, Vol. 2012, No. 1, p. 020028.

Avatar Partners, 2017, "Vuforia Model Targets Application in Aircraft Maintenance," available at vuforia.com/case-studies/avatar-partners.html.

ASTM, 2015, *ASTM E3023-15, Standard Practice for Probability of Detection Analysis for â Versus a Data*, ASTM International, West Conshohocken, PA.

Bainbridge, L., 1987, "Ironies of Automation," in *New Technology and Human Error*, J. Rasmussen, K. Duncan, and J. Leplat (eds.), John Wiley & Sons, Chichester, UK, pp. 271–283.

Bato, M.R., A. Hor, A. Rautureau, and C. Bes, 2017, "Implementation of a Robust Methodology to Obtain the Probability of Detection (POD) Curves in NDT: Integration of Human and Ergonomic Factors," Les Journées COFREND 2017, 30 June–1 July, Strasbourg, France.

Bertović, M., 2016a, "Human Factors in Non-Destructive Testing (NDT): Risks and Challenges of Mechanised NDT," Dissertation, Bundesanstalt für Materialforschung und -prüfung (BAM), Berlin, Germany.

Bertović, M., 2016b, "A Human Factors Perspective on the Use of Automated Aids in the Evaluation of NDT Data," *42nd Annual Review of Progress in Quantitative Nondestructive Evaluation, AIP Conference Proceedings*, Vol. 1706, No. 1, p. 020003.

Case, N., 2018, "How to Become a Centaur," *Journal of Design and Science*, doi: 10.21428.61b2215c.

Cowen, T., 2013, *Average Is Over: Powering America Beyond the Age of the Great Stagnation*, Penguin Group, New York, NY.

Dudenhoeffer, D.D., D.E. Holcomb, B.P. Hallbert, R.T. Wood, L.J. Bond, D.W. Miller, J.M. O'Hara, E.L. Quinn, H.E. Garcia, S.A. Arndt, and J. Naser, 2007, "Technology Roadmap on Instrumentation, Control, and Human-Machine Interface to Support DOE Advanced Nuclear Energy Programs," Report No. INL/EXT-06-11862, Idaho National Laboratory, Idaho Falls, ID.

Forsyth, D., J.C. Aldrin, and C.W. Magnuson, 2018, "Turning Nondestructive Testing Data into Useful Information," Aircraft Airworthiness & Sustainment Conference, 23–26 April, Jacksonville, FL.

Fuchs, P., T. Kröger, T. Dierig, and C.S. Garbe, 2019, "Generating Meaningful Synthetic Ground Truth for Pore Detection in Cast Aluminum Parts," *Proceedings of Conference on Industrial Computed Tomography* (iCT2019), Padua, Italy.

Fukushima, K., and S. Miyake, 1982, "Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition," *Competition and Cooperation in Neural Nets*, pp. 267–285, Springer-Verlag, Berlin/Heidelberg, Germany.

Gerbert, P., 2018, "AI and the 'Augmentation' Fallacy," *MIT Sloan Management Review*, available at sloanreview.mit.edu/article/ai-and-the-augmentation-fallacy/.

Hao, K., 2019, "When Algorithms Mess Up, the Nearest Human Gets the Blame," *MIT Technology Review*, available at technologyreview.com/s/613578/ai-algorithms-liability-human-blame/.

Hinton, G.E., S. Osindero, and Y.-W. Teh, 2006, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Computation*, Vol. 18, No. 7, pp. 1527–1554.

Jahanzaib, I., and J. Jasperneite, 2013, "Scalability of OPC-UA Down to the Chip Level Enables 'Internet of Things'," in *Proceedings of 11th IEEE International Conference on Industrial Informatics* (INDIN), Bochum, Germany, pp. 500–505.

Jordon, H., 2018, "AFRL Viewing Aircraft Inspections through the Lens of Technology," available at wpafb.af.mil/news/article-display/article/1603494/afrl-viewing-aircraft-inspections-through-the-lens-of-technology.

Kobryn, P., E. Tuegel, J. Zweber, and R. Kolonay, 2017, "Digital Thread and Twin for Systems Engineering: EMD to Disposal," 55th AIAA Aerospace Sciences Meeting, 9–13 January, Grapevine, TX.

LeCun, Y., Y. Bengio, and G. Hinton, 2015, "Deep Learning," *Nature*, Vol. 521, No. 7553, pp. 436–444.

Lewis-Kraus, G., 2016, "The Great A.I. Awakening," *The New York Times Magazine*, available at nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html.

Lindgren, E.A., 2017, "Opportunities for Nondestructive Evaluation: Quantitative Characterization," *Materials Evaluation*, Vol. 75, No. 7, p. 862–869.

Lindgren, E.A., J.R. Mandeville, M.J. Concordia, T.J. MacInnis, J.J. Abel, J.C. Aldrin, F. Spencer, D.B. Fritz, P. Christiansen, R.T. Mullis, and R. Waldbusser, 2005, "Probability of Detection Results and Deployment of the Inspection of the Vertical Leg of the C-130 Center Wing Beam/Spar Cap," 8th Joint DoD/FAA/NASA Conference on Aging Aircraft, 31 January–3 February, Palm Springs, CA.

Link, R., and N. Riess, 2018, "NDT 4.0 – Significance and Implications to NDT – Automated Magnetic Particle Testing as an Example," 12th European Conference on Non-Destructive Testing (ECNDT 2018), 11–15 June, Gothenburg, Sweden.

Meier, J., I. Tsalicoglou, and R. Mennicke, 2017, "The Future of NDT with Wireless Sensors, AI and IoT," 15th Asia Pacific Conference for Non-Destructive Testing, 13–17 November, Singapore, Singapore.

Meyendorf, N.G., L.J. Bond, J. Curtis-Beard, S. Heilmann, S. Pal, R. Schallert, H. Scholz, and C. Wunderlich, 2017a, "NDE 4.0—NDE for the 21st Century—the Internet of Things and Cyber-Physical Systems Will Revolutionize NDE," 15th Asia Pacific Conference for Non-Destructive Testing, 13–17 November, Singapore, Singapore.

Meyendorf, N.G., R. Schallert, S. Pal, and L.J. Bond, 2017b, "Using Remote NDE, Including External Experts in the Inspection Process, to Enhance Reliability and Address Today's NDE Challenges," 7th European-American Workshop on Reliability of NDE, 4–7 September, Potsdam, Germany.

Müller, C., M. Bertovic, D. Kanzler, T. Heckel, and R. Holstein, 2014, "Assessment of the Reliability of NDE: A Novel Insight on Influencing Factors on POD and Human Factors in an Organizational Context," *Proceedings of the 11th European Conference on Non-Destructive Testing* (ECNDT 2014), Prague, Czech Republic.

Müller, C., M. Golis, and T. Taylor, 2000, "Basic Ideas of the American-European Workshops 1997 in Berlin and 1999 in Boulder," *Proceedings of the 15th World Conference on Non-Destructive Testing* (WCNDT), Rome, Italy, pp. 1–7.

Olden, J.D., and D.A. Jackson, 2002, "Illuminating the 'Black Box': A Randomization Approach for Understanding Variable Contributions in Artificial Neural Networks," *Ecological Modelling*, Vol. 154, No. 1–2, pp. 135–150.

Oliver, N., T. Calvard, and K. Potočnik, 2017, "The Tragic Crash of Flight AF447 Shows the Unlikely but Catastrophic Consequences of Automation," *Harvard Business Review*, available at hbr.org/2017/09/the-tragic-crash-of-flight-af447-shows-the-unlikely-but-catastrophic-consequences-of-automation.

Rastogi, A., 2017, "Artificial Intelligence–Human Augmentation Is What's Here and Now," *Medium*, available at medium.com/reflections-by-ngp/artificial-intelligence-human-augmentation-is-whats-here-and-now-c5286978ace0.

Rice, S., and S. Winter, 2019, "The Boeing 737 Max Crisis Won't Stop the March of Airline Automation," *Quartz*, available at qz.com/1580078/the-boeing-737-max-crisis-wont-stop-airline-automation.

Rummel, W.D., 2010, "Nondestructive Inspection Reliability—History, Status and Future Path," *Proceedings of the 18th World Conference on Nondestructive Testing*, Durban, South Africa, pp. 16–20.

Sarter, N.B., D.D. Woods, and D.R. Billings, 1997, "Automation Surprises," *Handbook of Human Factors & Ergonomics*, 2nd ed., G. Salvendy (ed.), Wiley, New York, NY, pp. 1926–1943.

Scharre, P., 2016, "Centaur Warfighting: The False Choice of Humans Vs. Automation," *Temple International and Comparative Law Journal*, Vol. 30, pp. 151–165.

Sharp, T. D., J.M. Kesler, and U.M. Liggett, 2009, "Mining Inspection Data of Parts with Complex Shapes," *Proceedings of ASNT Annual Conference*, American Society for Nondestructive Testing, Columbus, Ohio.

Singh, R., 2019, "The Next Revolution in Nondestructive Testing and Evaluation: What and How?," *Materials Evaluation*, Vol. 77, No. 1, pp. 45–50.

Skagestad, P., 1993, "Thinking with Machines: Intelligence Augmentation, Evolutionary Epistemology, and Semiotic," *Journal of Social and Evolutionary Systems*, Vol. 16, No. 2, pp. 157–180.

Stapleton, A., 2017, "DARPA's Latest Take on AI Illuminates the Course of Intelligent Assistance," available at opusresearch.net/wordpress/2017/03/01/darpas-latest-take-on-ai-illuminates-the-course-of-intelligent-assistance/.

Taleb, N.N., 2007, "Black Swans and the Domains of Statistics," *The American Statistician*, Vol. 61, No. 3, pp. 198–200.

Taleb, N.N., 2012, *Antifragile: Things That Gain from Disorder*, Random House, New York, New York.

Tharp, D., 2017, "What Cyborg Chess Can Teach Us about the Future of Financial Planning," posted 24 July, available at kitces.com/blog/cyborg-chess-advisor-teach-about-future-financial-planning/.

US DOD, 2009, *MIL-HDBK-1823A*, "Nondestructive Evaluation System Reliability Assessment," US Department of Defense.

Vrana, J., K. Kadau, C. Amann, and D.U. Schnittstellen, 2018, "Non-Destructive Testing of Forgings on the Way to Industry 4.0," *Proceedings of ASNT Annual Conference*, American Society for Nondestructive Testing, Columbus, OH.

Wilson, J.H., and P.R. Daugherty, 2018, "Collaborative Intelligence: Humans and AI Are Joining Forces," *Harvard Business Review*, available at hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces.