Enough Fluff: Returning to Meaningful Perspectives on Automation

William N. Kaliardos
Office of NextGen
Federal Aviation Administration
Washington DC, USA
bill.kaliardos@faa.gov

***Abstract*—As society progresses towards increased automation in aviation—such as with Advanced Air Mobility and Unmanned Aircraft Systems—it is important to have a common understanding and perspective about automation among the many stakeholders, including aviation system designers, operators, maintainers, and regulatory authorities. Unfortunately, the discourse is hindered by misleading perspectives, assumptions, claims, and terminology.**

There are many examples. The term "automation" can be simply defined, but it is often confounded with "autonomous" and other descriptions of the function being automated, and further confounded by our subjective opinions on which functions are considered "advanced" or "intelligent". Automation is often discussed not as a tool that can be leveraged to achieve goals of the aviation community, but rather as a technocentric goal in itself. We often refer to automation as "an AI" (artificial intelligence) or a "team member", or other ways in which we anthropomorphize machines, yet do not clearly define functions for automated components of these desired systems. We argue that humans are prone to errors and that more automation therefore means fewer errors, without a fair balance that considers humans as valuable functional elements. We talk about operator trust as if the idea is unique to AI, when in fact the basic principles for human-automation interaction have not changed. We try to treat automation as a one-dimensional variable, such as with automation levels, but this hides important detail and has limited value in applications such as design, operations, and approvals of complex human-automation systems.

This paper identifies issues in recent automation and human-automation discourse, and provides clarifications and recommendations to improve progress towards the integration of increased automation in aviation systems.

Keywords—automation, autonomy, intelligent, human-automation, human-computer, human-machine, HMI, HCI, HSI, HAT, levels of automation, teams, teaming, artificial intelligence, machine learning, neurosymbolic, human error, function allocation

I. Introduction

The field of aviation may be on the cusp of a transformation in which automation is expected to be a key enabler. This includes artificial intelligence (AI), or at least its better-defined AI subset of machine learning (ML) algorithms, which have indeed introduced new opportunities in automation. While AI/ML has brought renewed attention to automation, it has also brought distracting and misleading information related to automation and human-automation integration, with far-reaching effects, even on the human side of the human-automation duality.

This paper addresses the topic of automation in aviation, including future "transformative" aviation systems such as Advanced Air Mobility (AAM) and Unmanned Aircraft Systems (UAS). It identifies misleading perspectives, terminology, claims, assumptions and other issues in recent automation discourse, and provides clarifications and recommendations in the following topic areas:

- Predicting automation capabilities
- Categories of automation
- Human-automation teaming
- AI and human factors
- Human and automation errors

The arguments presented are based primarily on prior (vs. new) research, and historical lessons. The intention of this work is to return to meaningful perspectives on automation, and to facilitate the appropriate integration of automation within future aviation systems.

II. Predicting Automation Capabilities

One of the most fundamental challenges in automation discourse is the simple question, "What can be automated?" This is like asking the question, "What can be modeled?" The famous response to that question is, "All models are wrong, but some are valuable". In a similar sense, functions "can" be automated, but whether automation will be valuable or appropriate is generally not answerable. That is, there is no general theoretical basis for determining which functions can be automated, or, in other words, for predicting automation capabilities.

Submitted for publication to the 41st Digital Avionics Systems Conference, September 18-22, 2022, Portsmouth, VA

The scope of automation considered here is for new or novel applications of automation for complex functions, such as cognitive tasks traditionally performed by humans. The conceptual choice whether or not to automate a function (or allow an operator to automate a function) is a complex design decision that involves many considerations, assumptions and uncertainties of hardware/software, operational environments, operator training, etc., perhaps most important of which is the function being automated. Even if we limit considerations to just functional performance, versus broader considerations like ethics, jobs, and human factors, the details and nuances of a function within its operating environment are unique, and the possible algorithms (means of automation) infinite.

A. Historical Lessons

Historically there have been many attempts to simplify the approach to determining what can be automated. In early research, [1] defined a list of functions that "machines are better at" or "man is better at" (MABA-MABA). There are heuristics like the "3 Ds": automate what is Dull, Dangerous, or Dirty. There are many more such lists and models for automating (particularly in comparison to humans), which become increasingly complex when moving away from simple heuristics, eventually being impractical to use. Such recipes for what should be automated fall short in application in part because they cannot capture the complex multivariable tradeoffs that are needed and the almost infinite ways humans and automation can work together [2].

There are also lessons to be learned from the field of AI. Early on, many claims were made about the potential of AI—notably symbolic AI such as Expert Systems—to achieve intelligence capabilities of human experts. One of the underlying assumptions was that expert knowledge could be elicited and codified in the form of symbolic rules. Despite evidence that human intelligence and expert knowledge in particular involves a tacit knowledge component that eludes declarative knowledge elicitation, critiques of the AI claims were often met not with respectful disagreement, but with derision. An early example is Hubert Dreyfus [3], who for many years endured attacks by the AI community, long before connectionist or nonsymbolic AI such as ML became practical. Most of Dreyfus' criticisms were eventually vindicated after years of failed demonstrations by the AI community, primarily those of Expert Systems, leading to the first "AI Winter" in the 1970s.

It may be that one of Dreyfus' arguments (human experts do not behave according to explicit rules) is as defensible as it can ever be regarding the ability of automation to achieve human-like intelligence. For example, explicit rules could offer an alternative method of achieving intelligence, and might serve to automate functions if the function is at least understood sufficiently well by developers—but who can precisely define limits to what is achievable? Just because we can't conceive of rules doesn't mean such rules can't be conceived or that such rules do not exist [4]. In general, the field of AI continues to provide new insights and opportunities, but its history also exemplifies the challenge in predicting what automation can or cannot do, and substantiating such predictions.

B. Demonstrations of Automation

Rather than theoretical proofs of what functions can be automated, we primarily depend on analogies to what has been demonstrated to work by whatever means available. That is not a theoretical basis; that is an observation of specific cases with specific algorithms, which can further be interpreted for potential application to similar cases. For example, we can now predict the near-term performance of future automated speech recognition systems (which had achieved impressive performance prior to AI) because of their extensive history of demonstrations and operational use. In the early attempts of speech recognition, such predictions had no substantive basis.

C. Safety-critical Automation

Even when specific cases of automation have been demonstrated, *safety-critical* applications present additional challenges, especially for very low-probability events. Furthermore, some events are a type called "unknown unknowns". That is, we can assume there are events that could occur during operation, but that we do not know specifically what they are. Consideration of unknown unknown operational events are safety and validation challenges, and an important aspect of resiliency. They are even more important as the functions being automated encroach on what have traditionally been performed by humans, partly due to the belief that humans can often better adapt to situations that were not explicitly trained, nor explicitly considered during design. Although that belief may be shifting with ML, for example, it remains that unknown unknowns are an important aspect of safety, and further limit our ability to understand automation's capabilities.

In summary, in the absence of precedence such as tests and demonstrations of specific functions in specific environments, one cannot definitively predict what automation can or cannot achieve. Perhaps the most we can say is that developers need sufficient understanding for symbolic algorithms, and developers need sufficient data for training ML algorithms—both of which say nothing about the nature of the function being automated. The devil is in the details of the algorithms and intended functions and operational environments, and many other variables. The consequence is that specific tests and demonstrations should be relied upon over general claims about automation capabilities. Another consequence is that strategic plans and roadmaps of aviation capabilities should depend on automation only if automation has at least been sufficiently demonstrated. Automation should generally be viewed as a potential means to a capability, and not a goal in itself.

III. AUTOMATION CATEGORIES

Another challenge with automation is in how we simplify the complex reality of aviation systems and the myriad of ways that automation can be integrated into aviation system operations. Automation categories, such as levels, are one way to simplify this complexity through a perceived reduction in dimensionality, such as in categories of common attributes. While there can be some value to an automation taxonomy, their historical use has not established precedence in application to complex system design, operations, and approvals.

A. Definitions and Assumptions

For discussions like this, some definitions and assumptions are needed. First, "automate" is defined here as simply the physical realization (implementation) of a function in hardware and software. A function might be abstract and described at various levels of detail, but implementing an intended function in hardware and software vs. humans (i.e., to automate a function) is a simple, objective determination. "Automation" or "automate" should not be confused with "autonomous". The term "autonomous" is fraught with relative and subjective attributes, and is often a distraction in the context of defining functions. "Automate" refers to a binary variable whose states are human (fully manual) or fully automated, and its state is of limited value without the additional consideration of functions and subfunctions.

Second, it is the *function* that describes *what* is automated, versus how it is automated. Although there is no limit to the number of functions that may be defined, or the degree to which a function can be decomposed into subfunctions, this enormous variability can be categorized into simple pieces and combined with the binary human-automation variable to create what is known as an "automation category", sometimes as "levels of automation" or "degrees of automation". These have taken many forms, and with various purposes. One early example is a two-dimensional taxonomy in which the function dimension is categorized as *information acquisition, information analysis, decision and action selection,* and *action implementation* [5].

For automation taxonomy purposes, it is reasonable to consider at least two dimensions together: a) The human-automation dimension, and b) the function dimension.

B. Potential Issues with Automation Categories

Automation categories have been used and misused over the decades. There can be many benefits to automation categories, such as to allow different perspectives and insights, or as a way to structure research. But there are also well-known issues, often stemming from their misapplication. Many of the pros and cons have been articulated cogently, as in [6], which identifies shortcomings such as oversimplification and discretization of a continuous and complex automation variable space. The following discussion builds upon these prior arguments, with a focus on the application of automation categories for normative versus descriptive purposes, and in the context of design and operation of aircraft, and their design/operational approvals.

C. Importance of Functional Detail

One well-known problem in the use of automation categories is the issue of insufficient functional detail. For example, one might describe an entire aircraft as having a certain degree of automation, without any description of the aircraft functions. While this is theoretically possible (perhaps in fully automated or fully manual aircraft), it implies that all the functions on the aircraft are automated similarly. For reasonably complex systems, and in the context of the human-centered aviation system that exists today, the myriad of functions combined with the myriad of ways those functions can be automated results in a combinatorial explosion along the two dimensions defined (the human-automation dimension combined with the functional dimension). "How is that aircraft automated?" does not have a simple answer.

D. Categories in Aircraft Design Approvals and Operations

Consider how manned aircraft are approved today, such as through a Type Design approval by the Federal Aviation Administration (FAA). The applicant, such as an aircraft manufacturer, has to comply with a host of federal regulations which are applied to potentially hundreds of defined functions and subfunctions. The intended functions are analyzed thoroughly by a number of technical specialists, driven primarily by safety considerations. That analysis strongly depends on the details of a function as well as how that function is implemented, and many other considerations, including assumptions of human operator roles. Automation categories are not applied because they hide important detail, and are not useful for the approval tasks. Function details and automation details (vs categories) are important in the analysis for regulatory compliance (e.g., design approvals), and need to be examined together versus independently. Furthermore, high degrees of integration at the aircraft level necessitate an analysis of how functions interact with other functions on the aircraft and throughout other parts of the aviation system. The point is that this aircraft design approval process today accommodates the necessary scrutiny to address automation integration complexity, and is likely incommensurate with typical categories of automation that inherently hide detail.

Automation categories in operational aspects, including airspace procedures, flight crew standard operating procedures, and flight crew training, face similar issues. For example, in commercial operations, a systematic, data-driven analysis of the operational use of flight path management systems found that automation levels were sometimes attempted to be used by airline operators, but many operators needed to revise their policy to reflect actual automation use [7]. The rigid, simplistic, and linear structure of automation levels does not stand up to the complexity of actual airline operations.

E. Lessons from Automobiles

In the apparent race toward automation, road vehicles perhaps are leading the way, and might provide some lessons regarding automation categories.

SAE J3016 [8] is a standard that defines six human-automation "levels", from Level 0 (fully manual) to Level 5 (fully automated).

The SAE levels are based on 5 function variables related to driving tasks, paraphrased for this discussion as follows:

- Long – Longitudinal control (e.g., speed)
- Lat – Lateral control (i.e., steering)
- Event – Object and event detection and response (e.g., collision avoidance)

- Fallback – Contingency mode of operation, such as after a system failure or change in the operational environment
- Environment – Assessing the operational domain (e.g., weather, terrain, day/night) with respect to the limitations of a function (e.g., lateral control)

Even though the SAE levels are designed to be "descriptive and informative, rather than normative", they illustrate the degree of simplification typically needed for an automation taxonomy. Conceivably, Long, Lat, and Event can be considered the primary driving functions, with each having a normal and a Fallback mode: a 3x2 function space. Additionally, each function within that 3x2 space can involve Environment to evaluate the conditions appropriate to the functions (Environment is described in [8] as a system vs human attribute, but in this discussion is broadened to include humans). This 3x2x2 space (12 states) does not even address further distinctions within each function, such as decisions vs actions, which, despite the low granularity, brings the number of functional states to 24. Of these 24 states, at a vehicle level, one can also consider the many possible *combinations* of functions, and additionally the combinations of human/automation allocation. Actual designs can introduce additional layers of how humans and automation might combine, including the possibility of those combinations dynamically changing, such as from mode selections. Although not every combination might be relevant in a taxonomy, the point is that even in the case of a small set of functions the possible combinations within a vehicle quickly becomes large, and difficult to capture in simple yet meaningful categories.

Yet, SAE defined only 6 levels. This is achieved through selection of specific combinations of functions and automation for each level, combining both dimensions into one, and with unclear rationale other than some progression of more automation. For example, Level 1 covers the specific combination of Long OR Lat, while Level 2 covers Long AND Lat. While the precise rationale of the taxonomy is not described, the result is indeed a simple scheme that is perhaps appropriate for the intended descriptive purpose, and which "make it possible for a range of engineering disciplines to talk about the next generation of cars" [9]. But the resulting loss of dimensionality necessitates numerous explanatory notes and examples in the SAE standard to explain the levels, revealing some of the hidden true complexity. Similar issues are inherent to other automation category schemes as well, suggesting that linear or even monotonic descriptions of automation are limited and could be misused to characterize a multidimensional space of functions and means of automation.

F. Aircraft Levels of Automation

In recent years, a new set of automation categories began to emerge within the aviation community, such as in Urban Air Mobility (UAM). (The general concept has been broadened in scope and named AAM). For example, the FAA UAM Concept of Operations [10] defines the following categories to describe "the evolution of aircraft automation":

- Human-within-the-Loop (HWTL): Human is always in direct control of the automation (systems)
- Human-on-the-Loop (HOTL): Human has supervisory control of the automation (systems), and actively monitors the systems and can take full control when required or desired
- Human-over-the-Loop (HOVTL): Human is informed, or engaged, by the automation (systems) to take action. Human passively monitors the systems and is informed by automation if, and what, action is required. Human is engaged by the automation either for exceptions that are not reconcilable or as part of rule set escalation

Table I provides a simplified breakdown of these UAM categories to highlight the subfunctions of human control and monitoring within each category, and to easily compare key attributes.

Is this taxonomy helpful? Given that the purpose is descriptive, the levels do convey a possible "evolution of automation", which is the stated intent. Are there possible issues with this taxonomy? There are a number of potential concerns, depending on more specific purposes:

1) Unclear boundaries on a continuum

As is often the case with automation levels, the issues tend to lie in the middle. In this case, HOTL's "supervisory control" can cover just about anything with respect to the degree of directness of control and the degree of monitoring, both of which are essentially on a continuum, with HWTL and HOVTL representing the extreme cases (end points) on the continuum. As one example, HOTL allows for humans to take full control when desired, in which case direct control (HWTL) is an option during operation. That sounds like how many aircraft operate currently: pilots often engage flight path automation, and take more direct control when necessary. On the other end of the spectrum, definitions of supervisory control typically include cases in which automation informs humans when to intervene (whatever the reason), so HOVTL also sounds very familiar in current aircraft. Does HOVTL differ from HOTL by precluding full control by humans, but still allowing some control? In HOVTL, can humans intervene when they desire, or are they restricted to wait until automation dictates how to intervene? What is the boundary between active and passive monitoring? How do alerts fit in—they inform humans whether an action is required, and sometimes what action is needed, and seem to potentially exist in all three categories. The definitions of HWTL, HOTL, and HOVTL illustrate a progression along a human control/monitoring continuum, but their lack of clear boundaries and overly

TABLE I. BREAKDOWN OF UAM AUTOMATION LEVELS

Human Role	**Levels**		
	HWTL	**HOTL**	**HOVTL**
Human Control	Always in direct control	Supervisory: can take full control when required or desired	When automation informs human if and how action is required
Human Monitoring	Active	Active	Passive

constraining definitions may present challenges to real world categorization.

2) Aircraft versus aircraft functions

An aircraft can be characterized by many functions, so, logically, an aircraft as a whole should not fall into any one category unless every function on that aircraft falls into that same category. Typically, aircraft use automation in a variety of ways, depending on the function and many other factors. Considering the possible number of combinations of functions and human-automation implementations, it may not be reasonable to characterize an aircraft in one of only three categories.

3) Coarse and linear categories

The categories appear along a single dimension, and can be interpreted as "levels". However, as with the SAE road vehicle levels discussed earlier, the human-automation dimension and the function dimension are combined into one. Each dimension is furthermore about as coarse as automation categories could possibly be, with the function dimension being restricted to generic observer/controller aspects (e.g., active vs passive monitoring). This low granularity and linearization can seem beneficial in that it allows for very simple perspective of complex systems and operations. However, such a reduction in dimensionality suggests a very limited practical use.

Without an understanding of these concerns, there is a risk of applying automation categories inappropriately. Recently the UAS Beyond Visual Line of Sight (BVLOS) Aviation Rulemaking Committee report [11] provided recommendations to the FAA, and identified categories of operations called "Automated Flight Rules" (AFR). AFRs establish very coarse and linear "autonomous levels" of operations for aircraft versus individual functions. Each category has associated human roles/qualifications, automation functions for flight controls, and aircraft equipage requirements (to-be-determined), in which the automation levels mirror the UAM categories defined above. The report appears to recommend that automation levels should dictate or at least map directly to pilot qualification and equipment requirements.

Although the ARC report is advisory, and an early step in the conversation about UAS BVLOS regulation, its definition of automation levels and mapping to pilot qualifications and equipment requirements implies a potential to apply automation levels prescriptively towards technical decision making related to safety. As mentioned, aircraft can be described by a large number of functions, each of which can potentially use automation in many ways that are not always represented sufficiently in coarse, linear, predefined human-automation categories. Prescribing pilot qualifications or other design or operational requirements depend on the details of precisely how humans and automation interact, as well as the details of the functions being automated, and many other co-dependent variables. The level of detail needs to be commensurate with the analysis or decisions, so attempts to streamline these details through coarse human-automation categories might have little or no precedence from past manned aircraft, and limited value in characterizing or prescribing UAS or other aviation systems and operations.

IV. Human-Automation Teaming

Another challenge in the discussion of automation is related to systems that are characterized as human-like in behavior. As mentioned, general claims about automation's potential capabilities are difficult to substantiate. Even so, in the context of engineering it can be a distraction to describe systems (e.g., hardware/software functions or requirements) as anthropomorphisms that are subjective or otherwise ill-defined. Examples include descriptions such as "intelligent", "autonomous", and "aware". While not all anthropomorphic terminology is inappropriate, nor is humanness only achieved through automation [12], such terminology likely provides little value in an engineering sense unless further translated into what can be evaluated against well-defined criteria. Even those in the field of AI typically wish to escape their label, in part because there is no universally accepted definition of "intelligence", and because AI becomes a moving target, relative to current capabilities.

In system design and approval, it is critical that all functions are described unambiguously and at appropriate levels of detail.

A. Automation as a Team Member

Anthropomorphism is particularly rampant when applied to "human-automation teaming", or HAT. The intent of HAT concepts is for automation to behave less like a tool that is controlled by humans, and more like a proactive human team member (perhaps even called "an AI"). Moving beyond the traditional human-computer interaction (HCI) principles, the defining characteristics of HAT often include:

- Mutual coordination of tasks
- Pursuit of shared goals:
- Shared situation awareness
- Shared understanding
- Bidirectional communication of intent, tasks, and actions/decisions
- Mutual trust

Already we can observe that the name, "teaming," and many of the HAT characteristics are inherently human-like. On the surface, they sound like worthy design aspirations towards intelligent systems. But in the context of design and approval there are significant ambiguities. For example, what does it mean for a *system* to "understand" or to "trust" or be "aware"? Reference [13] describes the automation component of HAT as a perception by the operator.

There is a difference between what might be *perceived* as teaming in an operational sense, and the functional definitions that characterize HAT systems for design and design approval. These differences are not new; the HAT attributes might sound familiar to those who follow AI history. Even though the HAT characteristics often are identified as characteristics of high performing *human* teams, that does not mean it is reasonable to make the leap to define these as functions of automation for human-automation teams, nor reasonable to avoid translating these into proper functional descriptions.

B. Anthropomorphic Terminology in Aircraft Approvals

The FAA Aircraft Certification Service provides aircraft design approvals based on regulations that apply to systems, none of which attribute human-like terminology to the systems versus the effects on the humans using those systems. An example is an FAA regulation on controls [14], which states, "Each cockpit control must be located to provide convenient operation and to prevent confusion and inadvertent operation". The terms "convenient" and "confusion" apply to the display's effects on human pilots, and therefore need to be evaluated in that sense. Such regulations are often purposely general in part to allow for various means of compliance, but they need to be commonly understood, often with the aid of guidance such as advisory circulars. Furthermore, compliance determination often additionally requires supporting data, and evaluations from pilots, including professional judgment of test pilots. Regulations that refer to system effects on humans present unique challenges, and can involve a suite of guidance and methods to determine regulation compliance effectively and consistently. In this example, whether a control is "convenient" or whether it prevents "confusion" is challenging in part because that determination requires a degree of subjective judgment in estimating effects on the intended pilot population. In comparison, a HAT function defined by human-like terminology (e.g., "mutual trust") would likely be more challenging, and need additional unambiguous function definitions, and possibly new alternative regulations and methods of evaluation.

As another example, the FAA dissuades the term "situation awareness" as intended functions of systems. An FAA advisory circular on aircraft electronic displays states, *"General and/or ambiguous intended function descriptions are not acceptable (for example, a function described only as "situation awareness"). Some displays may be intended to be used for situation awareness, but that term needs to be clarified or qualified to explain what type of specific situation awareness will be provided"* [15]. As a specific example, Traffic Alert and Collision Avoidance Systems (TCAS) may include traffic displays along with directive alerting systems. The intended function of the traffic display might be considered to be *situation awareness* by many, but is specifically defined as how the information should be used by pilots in their decisions and actions—such as to assist pilots in the visual acquisition of traffic out the window, to provide pilots with confidence in proper system operation, and to prepare for a maneuver if certain alerts are issued.

Situation awareness exemplifies the need to replace or supplement functional descriptions with appropriate detail, even when applied to system effects on users. Without further interpretations and detail of HAT (e.g., of "shared situation awareness", "shared understanding"), it is reasonable to suggest that applying such anthropomorphic descriptions to systems (versus users) will remain incompatible with current FAA approval processes.

HAT may be on the path to repeat similar mistakes from the past decades of AI development in emulating human behavior. There will always be value in pushing the limits of automation within human-automation interaction. This progress should not distract from clearly defining HAT system functions for design and design approval purposes, separately from how they might be perceived as human-like by operators.

V. AI and Human Factors

What, then, is the relevance of AI to the field of human factors? It should be accepted that human factors is critical throughout aviation systems design, operations, approvals, maintenance, etc., especially with the introduction of new systems and automation. New or novel automated functions introduce unknown rippling effects (to other systems and humans). A key driver of the need for human factors scrutiny is not AI, per se, but *new automation capabilities (or functions)*. Whether automating functions is achieved through AI or traditional algorithmic means is often not relevant to human factors—at least not in the context of operations. This section identifies potential points of confusion surrounding the impacts of AI on human factors, with a focus on ML.

A. Operationally Relevant Information

First, an assumption is made here that humans need to exchange information with an automated system, and for this discussion, such information is limited to be what is "operationally relevant". Operationally relevant information is a term used here to, for example, distinguish detailed algorithm information from that which is directly relevant and appropriate to supporting and operator's task.

Operationally relevant behavior is a broader term, and an important aspect of an FAA regulation and its advisory circular that address the operational use of aircraft systems by flight crews [16][17], which uses the phrase "operationally relevant behavior" towards information, the system's operational logic, and other aspects of a system—not the detailed logic of the software. The advisory circular states that operationally relevant behavior "distinguishes between the system behavior as perceived by the flightcrew and the functional logic of the systems flightcrews operate."

During *operations*, much of what goes on "inside the box" (or, to use a vehicle analogy, "under the hood") is not operationally relevant. Most algorithmic details of automation, including whether it involves AI, or what type of AI, is typically not relevant to operators. For example, when one uses speech recognition on a smartphone, they don't know if AI was used—those details are under the hood. The operationally relevant aspects are overall speech performance such as accuracy, speed, etc. Sometimes, as in decision support tools, understanding the rationale or logic behind an automated decision recommendation might be included in the set of information that is operationally relevant, and some forms of AI are black boxes in that they are inherently limited in their ability to provide certain information. The idea of operational relevance is important to help understand the relationship between AI and human factors.

B. Machine Learning

One class of AI particularly relevant to this discussion is ML. ML considerations are driving many recent AI capability and AI interaction discussions. ML is characterized by algorithms such as neural networks whose algorithms are fully determined through training on data sets (this discussion assumes the algorithm does not further change during operations). There are many variations, such as if or how humans are part of training. However, the key aspect of ML for this discussion is that automated functions are not achieved through explicit, human-understandable rules and symbols, as is the case with traditional algorithms and symbolic AI. ML therefore has profound consequences on the means by which automation functionality can be achieved. In particular, it circumvents the need for designers to understand, for example, specifically "how" decisions are made, and instead shifts design towards how to learn, and understanding "what" data are appropriate for learning. ML also avoids many of the AI criticisms by Dreyfus [3], which at the time was primarily focused on symbolic AI (and claimed that human experts, opposed to novices, cannot articulate their knowledge via rules because their expertise is not based upon rules).

C. Two Types of Information Processing

In some sense, ML has similarities to human judgment and intuition, characterized by the absence of explicit rules and lack of structure, while symbolic AI and traditional algorithms have similarities to human analytical and deliberative thinking, characterized by the presence of explicit rules and structure. These two very different means of information processing, whether through human or automation, have been identified in various contexts, such as in [18–21], and can provide insights into the tradeoffs between different types of human and automated processes. However, they are not intended to prescribe what functions can be automated. The tradeoffs might suggest that ML is not a silver bullet for AI, just as Expert Systems were not a silver bullet for symbolic AI, and furthermore that perhaps some "neurosymbolic" or "semi-structured" combination might be the most appropriate—even though the type of AI is typically not directly relevant to the operator.

D. Relevance of AI to the Operator

With that background, three arguments are presented to distinguish the effects of AI/ML vs. other forms of automation. First, ML offers very different opportunities in realizing automated functions because of the shift in design approaches, from explicit rules to data sets. However, this fundamental shift does not introduce new automation requirements related to ethics, bias, responsibility, and other characteristics of human decisions. It is understandable that automated functions formerly performed by humans might place a lens on human behaviors and attributes, but these qualities can be important considerations for automation in general, and are not specific to ML. ML merely has highlighted them because they are inherent in training data (involving humans) vs explicitly coded.

Secondly, human-automation *interaction* considerations, such as human trust of automation, are also not specific to the type of AI or other algorithm. Trust in automation and reliance on automation are important topics that have existed since automation research has existed, and are not unique to AI or ML. AI can influence the functions that are automated, but it is the operationally relevant behavior of automation (e.g., performance), including what is communicated at the human-automation interface, that largely influences operator trust.

Thirdly, one aspect of ML that is specifically relevant to operational human factors impacts is explainability. As mentioned, ML has a limitation inherent to the absence of symbolic algorithms, similarly to the limitation inherent with human expert knowledge. Explainability pertains to the ability to explain how or why decisions or other automation outputs occurred, and can therefore limit the transparency of ML automation. This can in turn limit operationally relevant information, such as that which describes the logic and rationale behind decisions and control actions.

When the above points are not clarified, there is a tendency to use "AI" and "automation" or "autonomy" interchangeably. As an example, the following are recommendations from a recent report on Human-AI Teaming [22], with the same key word omitted in each statement:

> *"While it is assumed that human-AI teams will be more effective than either humans or AI systems operating alone, …. this will not be the case unless humans can:*
>
> *(1) understand and predict the behaviors of the _____ system*
>
> *(2) develop appropriate trust relationships with the _____ system*
>
> *(3) make accurate decisions based on input from the _____ system*
>
> *(4) exert control over the _____ system in a timely and appropriate manner"*

To fill in the blanks, is the correct word a) automation, or b) AI? The report used "AI". However, either could reasonably apply, which is the point here: the statements are equally true with "automation" or "AI." They all prescribe well-known, desired attributes of human-automation interaction, so in that sense nothing new is being stated, yet it leads one to believe that these statements may not generally apply to other forms of automation. While the context of the report is AI, and the intention may be to call attention to the continued importance of human factors in the new age of AI, restating traditional automation requirements as AI requirements can be misleading. "AI" refers to a means of automation, and should not replace "automation" unless AI brings specific considerations to human interaction, such as explainability.

VI. Human and Automation Error

Finally, human vs automation "error" is sometimes discussed in a misleading way. It is common to hear generalizations such as "human error is a causal factor in 80% of aviation accidents and incidents". These generalizations are then further manipulated through an illogical claim that these errors can largely be eliminated if humans are replaced with automation.

That claim is illogical for numerous reasons, such as:

- Automation brings with it other forms of system errors , failures, and behaviors, which typically are not mentioned.
- Human errors are not eliminated when human tasks are replaced by automation, but, rather, shift, often in ways in which errors are more consequential [2]. Overall, changes in automation alter the overall human-automation interactions in unpredictable ways.
- Human contributions to safety (vs. human error) outside of safety events leverage resilient behaviors like adaptation [23], but are not well characterized in part because of their subtle roles normal and near-normal operations.
- "Errors" can be defined in many ways, and a binary definition is a simplification that is not necessarily appropriate for an analysis.
- Many aircraft incidents and accidents involve pilot challenges in understanding the situation, such as diagnosing automated system failure. More sophisticated automation may add complexity and exacerbate this challenge.

The claim of "eliminating" human errors through automation substitution is fraught with over-simplifications and misguided assumptions. For reasonably complex systems and operations it is challenging to perform a balanced "apples to apples" comparison of errors, and human-automation performance in general, without a detailed analysis that considers not only errors, but all relevant behaviors of humans and automated systems that affect safety. The challenges include identifying what is lost from human cognitive skills that are not well characterized yet important for continuously contributing to safety by avoiding accidents and incidents in the first place.

VII. Summary and Recommendations

Automation is expected to be a key enabler of transformative aviation systems such as AAM and UAS. In particular, advances in AI have fueled bold predictions of automation capabilities, but also misunderstandings in recent automation and human-automation discourse. In this work, five topic areas were addressed, with the following summary of some of the key findings and recommendations:

- **Predicting Automation Capabilities**. Since it is not theoretically possible to predict automation capabilities in a general sense, claims or plans about future aviation advancements enabled by automation might be misguided. Recommendation: Capability roadmaps should depend on automation only if automation has at least been sufficiently demonstrated. Automation should generally be viewed as a potential means to a capability, and not a goal in itself.
- **Automation Categories**. Automation categories such as "levels of automation" have been useful for some purposes such as research, but their inherent simplification of complex systems may limit their applicability in practice. Recommendation: Automation categories should be used cautiously for characterizing aviation systems and operations, especially for prescriptive purposes such as driving safety requirements.
- **Human-automation Teaming.** HAT concepts are often based on anthropomorphic functional descriptions of automated systems that are ambiguous. Recommendation: Automated system functions within HAT should be defined objectively and with specificity for design and design approval purposes, separately from how they might be perceived as human-like by operators.
- **AI and Human Factors.** AI brings many new opportunities for achieving automation capabilities, but human factors is often not uniquely impacted by AI *per se*, versus by other forms of automation. Traditional human-automation principles still apply in the new age of AI, and human factors application remains critical in design and operations, including approvals. Recommendation: In human factors research and application, the term "AI" should not replace "automation" unless AI brings specific considerations to human interaction, such as explainability.
- **Human and Automation Errors**. As automation capabilities change, the claim of replacing humans with automation and thereby eliminating human error is flawed. Recommendation: When automation is altered, errors, information, tasks, etc., can also be altered in unpredictable ways throughout the human-automation system; the analysis of these changes should be commensurate with the complexity of the system and operations.

This work has attempted to provide a logical basis for the findings and recommendations, based upon prior research historical lessons. One of the emergent themes is the importance of lessons from decades of AI progress and human-automation integration. The lessons should continue to stand firmly, even if quietly among the clamor of recent AI progress and the rush towards the future. It is hoped that this work will raise awareness and improve discourse towards more meaningful perspectives on automation in future aviation systems.

Acknowledgment

I would like to thank Dr. Divya Chandra of the U.S. Department of Transportation Volpe Center for reviewing this paper. The views expressed herein are those of the author and do not necessarily reflect the views of the Federal Aviation Administration.

References

[1] P. Fitts, "Some basic questions in designing an air-navigation and traffic-control system," in *Human engineering for an effective air navigation and traffic control system*, Washington, DC: National Research Council, ch. III, pp. 5-11, 1951.

[2] S. Dekker and D. Woods, “MABA-MABA or Abracadabra? Progress on human–automation co-ordination,” . *Cognition Tech Work*, vol. 4, pp. 240–244, Nov. 2002.

[3] H. Dreyfus, “Alchemy and Artificial Intelligence,” RAND Corporation, Santa Monica, CA, 1965, [Online]. Available: https://www.rand.org/pubs/papers/P3244.html

[4] S. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, Englewood Cliffs, N.J: Prentice Hall, 1995.

[5] C. Wickens, A. Mavor, R., Parasuraman, R., and J. McGee, (Eds.) *The future of air traffic control: Human operators and automation*, Washington, DC: National Academy Press, 1998.

[6] E. Roth and A. Pritchett (Ed.), “Special issue: Advancing models of human-automation interaction,” *Journal of Cognitive Engineering and Decision Making*, vol. 12, pp. 3–94, 2018.

[7] Federal Aviation Administration, “Operational use of flight path management systems: final report of the performance-based operations Aviation Rulemaking Committee/Commercial Aviation Safety Team Flight Deck Automation Working Group”, Sep. 2013.

[8] *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles*, Technical Report No. J3016, SAE International, Warrendale, PA, 2021.

[9] J.D. Lee, “Perspectives on automotive automation and autonomy”, *Journal of Cognitive Engineering and Decision Making*, vol. 12, pp. 53–57, 2018.

[10] Federal Aviation Administration, “Urban Air Mobility Concept of Operations, Version 1.0”, 2020.

[11] *Unmanned Aircraft Systems Beyond Visual Line-of-Sight Aviation Rulemaking Committee Final Report*, Mar 2022.

[12] J.B. Lyons, K. Sycara, M. Lewis, and A. Capiola, (2021) “Human–Autonomy teaming: definitions, debates, and directions,” *Front Psychol.*, 2021.

[13] K. Wynne and J.B. Lyons, “An integrative model of autonomous agent teammate-likeness,” *Theoret. Issues Ergonom. Sci.* vol. 19, pp. 353–374, Feb. 2018.

[14] Federal Aviation Administration, 14 C.F.R. §25.777(a), “Cockpit controls”.

[15] Federal Aviation Administration, Advisory Circular 25-11B, “Electronic flight displays,” 2014.

[16] Federal Aviation Administration, 14 C.F.R, §25.1302, “Installed systems and equipment for use by the flightcrew”.

[17] Federal Aviation Administration, Advisory Circular 25.1302-1, “Installed systems and equipment for use by the flightcrew,” 2013.

[18] H. L. Dreyfus and S.E. Dreyfus, *Mind Over Machine*, The Free Press, New York, 1986.

[19] K.R. Hammond, R.M. Hamm, J. Grassia, and T. Pearson, “Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment,” *IEEE Trans. Sys., Man, and Cybernetics*, vol. 17, no. 5, pp. 753-70, 1987.

[20] W. Kaliardos, “Semi-structured decision processes: A conceptual framework for understanding human-automation systems,” Ph.D. dissertation, Massachusetts Institute of Technology, 1999.

[21] D. Kahneman, *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 2011.

[22] *Human-AI Teaming: State of the Art and Research Needs*, Washington DC: The National Academies Press, National Academies of Sciences, Engineering, and Medicine, 2021.

[23] D.J. Provan, D.D. Woods, S.W.A. Dekker, A.J. Rae, “Safety II professionals: How resilience engineering can transform safety practice, Reliability Engineering & System Safety,” Volume 195, Mar 2020.