

Received May 26, 2018, accepted June 29, 2018, date of publication July 5, 2018, date of current version July 30, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2853560

Common Metrics to Benchmark Human-Machine Teams (HMT): A Review

PRAVEEN DAMACHARLA¹, (Graduate Student Member, IEEE),
AHMAD Y. JAVAID¹, (Member, IEEE), **JENNIE J. GALLIMORE²**,
AND VIJAY K. DEVABHAKTUNI¹, (Senior Member, IEEE)

¹Department of Electrical Engineering and Computer Science, The University of Toledo, OH 43606, USA

²College of Technology, Architecture and Applied Engineering, Bowling Green State University, Bowling Green, OH 43402, USA

Corresponding author: Ahmad Y. Javaid (ahmad.javaidd@utoledo.edu)

This work was supported in part by the Paul A. Hotmer Cybersecurity and Teaming Research (CSTAR) Lab, Electrical Engineering and Computer Science Department, University of Toledo, and in part by the Round 1 Award—Improving Healthcare Training and Decision Making Through LVC from the Ohio Federal Research Jobs Commission through Ohio Federal Research Network.

ABSTRACT A significant amount of work is invested in human-machine teaming (HMT) across multiple fields. Accurately and effectively measuring system performance of an HMT is crucial for moving the design of these systems forward. Metrics are the enabling tools to devise a benchmark in any system and serve as an evaluation platform for assessing the performance, along with the verification and validation, of a system. Currently, there is no agreed-upon set of benchmark metrics for developing HMT systems. Therefore, identification and classification of common metrics are imperative to create a benchmark in the HMT field. The key focus of this review is to conduct a detailed survey aimed at identification of metrics employed in different segments of HMT and to determine the common metrics that can be used in the future to benchmark HMTs. We have organized this review as follows: identification of metrics used in HMTs until now, and classification based on functionality and measuring techniques. Additionally, we have also attempted to analyze all the identified metrics in detail while classifying them as theoretical, applied, real-time, non-real-time, measurable, and observable metrics. We conclude this review with a detailed analysis of the identified common metrics along with their usage to benchmark HMTs.

INDEX TERMS Autonomous system, benchmarking, human factors, human-machine teaming (HMT), metrics, performance metrics, and robotics.

I. INTRODUCTION

The future of technology lies in human-machine collaboration rather than on a completely autonomous artificial intelligence (AI). Dr. Jim Overholt, senior scientist at the Air Force Research Lab (AFRL), stated, “The US Air Force Research Laboratory (AFRL) has no intention of completely replacing humans with unmanned autonomous systems” [1]. Therefore, to achieve the best results, a human-machine teaming or collaboration is the only choice we have, but such a teaming comes with its own set of challenges. We propose to define HMT as a combination of cognitive, computer, and data sciences; embedded systems; phenomenology; psychology; robotics; sociology and social psychology; speech-language pathology; and visualization, aimed at maximizing team performance in critical missions where a human and machine are sharing a common set of goals. Team members will share tasks, and the machine may

provide suggestions that can play a crucial role in team decision-making. Such a collaboration requires a two-way flow of information. Based on the above-proposed definition, to be deemed as an HMT, a team should contain at least one human and one machine/intelligent system. Perhaps the best example of practical use of an HMT can be attributed to a 2005 game of chess. In this game, two inexperienced chess players teamed up together with three PCs and won a chess competition against a group of supercomputers and grandmasters, which did not form a team. In this scenario, human team members were able to leverage the machine’s data mining and information processing capabilities based on their cognition skills [2]. Although machines have been used to assist humans for decades, these systems are not collaborative partners but are programmed for specific tasks [3]. The primary concern of HMT is effective integration of human and machine tasks so that the team collaboration optimizes

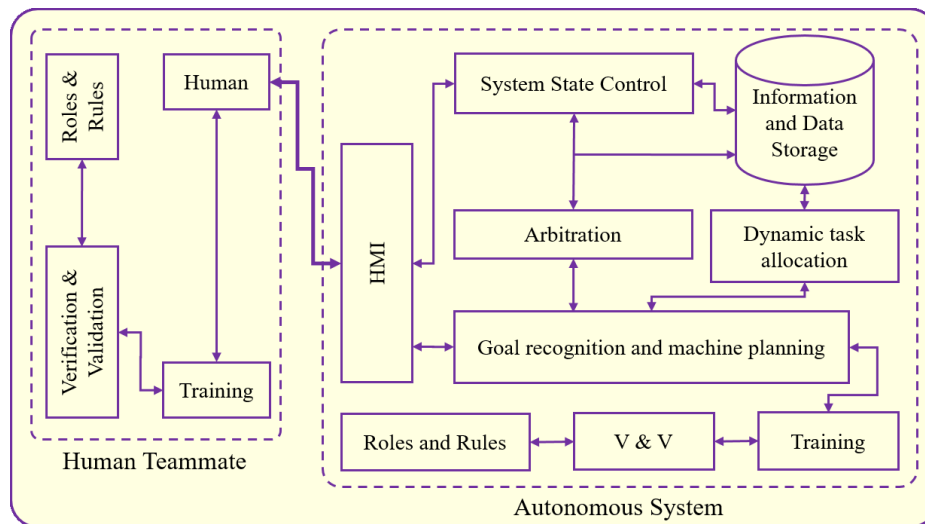


FIGURE 1. Generalized human machine team model.

the efficiency of critical tasks [4]–[8]. Making successful outcomes consistent and repeatable with high accuracy would also demonstrate an effective HMT that is possible only through comprehensive studies.

A. HMT OVERVIEW

By analyzing various published works [8]–[21]], we identified six major HMT components, with architectures, interfaces, and metrics being the highly researched areas. We present brief definitions and examples of these components:

1) **ARCHITECTURES**: The founding principle of building an HMT architecture is to achieve an optimal machine assistance. Architecture is necessary to set boundaries, assign duties, and design interfaces to increase the team effectiveness. Through analysis of 19 published frameworks, we identified nine essential functional blocks for a generic HMT framework: human-machine interaction (HMI), information and data storage, system state control, arbitration, goal recognition and mission planning, dynamic task allocation, rules and roles, verification and validation (V&V), and training [21]–[35]. This is shown in figure 1.

2) **INTERFACES**: Any focus over interface and interaction method will enable an effective human-machine communication. The association for computing machinery defines HMI as “a discipline concerned with the design, evaluation, and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them” [36]. The HMI can be divided into three principal components: the user, the interface, and the machine. Here, an interface is a device that typically encompasses both software and hardware to streamline an interaction between user and machine. Examples include a graphical user interface, web browsers, and various I/O devices [37]. Many published

studies that have classified and analyzed interfaces used in HMI are acknowledged in [38]–[40].

3) **METRICS**: Metrics are crucial measures to track, assess, and compare a process, task, or system with respect to performance, usability, efficiency, quality, and reliability as defined by the system performance goals. Metrics can also be used to evaluate the effectiveness of an HMT and its agents (human, machine, and team) on various levels.

4) **ROLES AND RULES**: Roles are defined as assumed or assigned responsibilities within a system, process, or task. On the other hand, rules are defined as a set of explicit regulations governing conduct in a situation or activity. By analyzing published work, we concluded that requisite and opportunistic are two categories of roles and rules. Implementing roles and rules in HMT helps generate a symbiotic human-machine ecosystem that will think as no human has ever thought and will process the data in a way that no machine ever processed [4], [31], [41]–[45].

5) **TEAM BUILDING**: According to the earlier works of researchers presented in [46], teams are defined not as just individual parts of machinery but they must be built together. In an HMT, one can build a systemic team with compatible team members. Through literature review, we identified that team development has two dimensions: (1) the task dimension consisting of forming, conflict resolution, norming, and performing, and (2) the interpersonal dimension consisting of dependency, conflict, cohesion, and interdependence [46]–[53].

6) **VERIFICATION AND VALIDATION (V&V)**: For a team to function optimally, features such as trust, cohesion, expectations, and motivation must be considered because of their effects on team performance. V&V is a crucial component of HMT that helps validate the team-building features mentioned above and thus gives key insights for optimizing

the team formation and performance. The V&V methods can be further classified into two groups based on their use: during mission and training [18], [54]–[59].

B. METRICS BACKGROUND

Although the foundations of HMT were laid at Defense Advanced Research Projects Agency in 2001 [10], it took five more years for the research community to identify a set of metrics that facilitates a well-organized structure of human-robot interactions (HRI). For various metrics, we found close but different descriptions of the same metric, primarily for various HMIs in human-robot or robot-only swarms [20], [60]. The research community uses metrics that are application and domain specific. For example, researchers in [61] developed an approach to define human supervisory control metrics while [62] has identified common metrics for HRI standardization. Researchers in [63] and [64] focused on developing false alarm metrics to analyze erroneous HRI. The robot performance evaluation metrics for understanding team effectiveness are defined in [65]. Researchers have also developed metrics from human-computer interaction (HCI) heuristics to aid information analysis in interactive visualizations [66]. This work made an active effort to define metrics for specific components of HMT, such as HCI, HRI, and architectures, whereas research on common metrics is limited.

Identifying common metrics will allow benchmarking of HMT designs, comparison of findings, and development of evaluation tools. The primary difficulty in defining common metrics is the diverse range of HMT applications. In this review, we focus on metrics for all three agents of HMT, for example, human, machine, and the team. The goals of this review paper are (1) identification and classification of metrics, (2) evaluation of the identified metrics to find common metrics, and (3) proposal of common metrics that can be used in future HMT benchmarking. The rest of the paper is structured as shown in figure 3.

II. METHOD

A. KEYWORDS AND DATABASES

To limit the scope of this study, we developed a set of keywords based on pertinent technological and scientific domains that focus on HMT. The HMTs investigated in this study account for one or more task-oriented mobile robots or software agents as machine team member(s) and at least one human as a team member. Further, the machines that take part in an HMT must belong to one of the following categories: unmanned aerial vehicles (UAVs), unmanned ground vehicles, AI robots, digital assistants, and cloud assistants, as shown in figure 2. The search was limited to HMT applications in target search and identification, navigation, ordinance disposal, geology, surveillance, and healthcare. The keywords used are listed in Table 1 and the databases utilized are as follows: IEEE Xplore, Science Direct

TABLE 1. Keywords used.

Primary Keywords	
Core Concepts	Keywords
Human Machine Teaming	Control metrics, interface metrics, synthetic assistant, synthetic mentor, intelligent assistant, rules and roles, symbiosis, verification and validation, measuring methods, physiological attributes
Human Machine Collaboration	Metrics, architecture, interface, team building, human factors, ergonomics, task, automation, shared control, symbiosis, physiological attributes
Human in Team	Team building, metrics, interface, human factors, human-robot collaboration, ergonomics, multirobot controls, shared control, physiological attributes
Machine in Team	Robot control, software agent, synthetic assistant, metrics, synthetic mentor, intelligent assistant, team building, interface, human factors, multirobot teams
Metrics	List of all identified metrics in Table IX + measuring methods
Secondary Keywords	
UAV, UGV, navigation, surveillance, healthcare, medical assistant, identification, ordinance disposal, geology	

(SCOPUS/ Elsevier), Defense Technical Information Center, SAGE Publications, and Google Scholar.



FIGURE 2. Trends in autonomous systems.

B. SELECTION CRITERIA

The following criteria were set to evaluate the articles found after a detailed search. Firstly, we tried to define the relevance of the article with our objectives/goals as follows:

- Discusses HMT or human-machine collaboration?
- Discusses one or more HMT components?
- Discusses metrics related to an HMT agent?
- Mentions or discusses core HMT concepts?

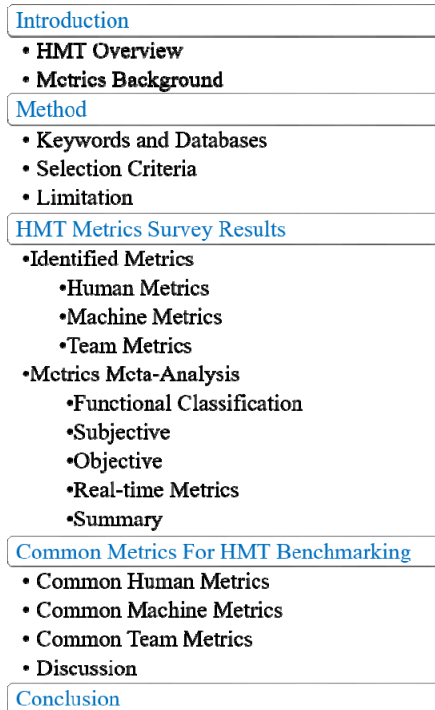


FIGURE 3. Review paper structure.

Articles that satisfied the above criteria were further filtered based on primary and secondary keywords used in specific sections. Further, we identified metrics that relate to teaming and HMT and conducted another refined search to obtain the most relevant literature. Out of hundreds of articles identified in the search process, a total of 188 articles were considered for the review.

C. LIMITATIONS

A key limitation of this review is the breadth of the review since the area of HMT is extensive and involves many fields of study. For the sake of this review, a limited number of primary articles are reviewed here ($n = 77$). Such a wide-range review poses a bigger challenge in terms of comprehensive coverage of various metrics and related research questions. Therefore, the review focuses on three agents of HMT that are worthy for an in-depth review. We selected the most relevant information available from the literature. Another limitation that entails establishing common metrics for all HMT types or benchmarking them on a single scale is the dependence on many factors such as application, and the number of agents.

III. HMT METRICS SURVEY RESULTS

In this section, we present a comprehensive and classified metric list for the three agents of HMT: human, machine, and team (or system). This strategy resulted in (1) an analysis that applies to a specific range of applications, and (2) the ability to assess the application specific HMT performance.

A. IDENTIFIED METRICS

1) HUMAN METRICS

This subsection identifies metrics that measure different human aspects such as system knowledge, performance, and efficiency that can be used to evaluate a human agent in an HMT. Most of the metrics we present in this section are well established by various scientific studies.

Situational awareness (SA) is measured by monitoring task progress and sensitivity to task dynamics during execution. The degree of *mental computation* estimates the amount of *cognitive workload* an operator manages to complete a task, for example, a task that requires object reference association in *working memory* or a user's cognitive abilities to perceive projections of the real-time environment [62], [67]. The accuracy of a *mental model* of an operator depends on interface comprehensiveness and simplicity in addition to control and compatibility a machine provides. *Attention allocation* measures the attention an operator pays to a team's mission and the operator's ability to assign strategies and priorities of tasks dynamically. The metric also considers an operator's degree of attention over multiple agents. It is measured using eye tracking, duration of eye fixations to an area of interest, and task completion rate, while *attention allocation efficiency* is measured using wait times [61], [68], [69]. *Intervention frequency* is the frequency with which an operator interacts with the machine [20]. As per literature, operators' intervention frequency is also known as *intervention rate* or *percentage requests*. *Stress* can be physical or mental. However, both may indicate the operator workload and are measured in two ways. First, researchers perform sample testing of humans' stress hormones, such as hypothalamic-pituitary-adrenaline, cortisol, and catecholamine, which are found in blood, saliva, and urine samples [70]. Second, researchers can perform a detrended fluctuation analysis of a human's heartbeat [71].

Human *safety metrics* involve evaluation of the risk posed to the human life while working near machines, for example, the location of the machine relative to the human. These mostly apply to applications in a high-risk environment such as threat neutralization. Human factor studies suggest that humans can establish the best cooperation with a machine through a 3D immersive environment [72]. In [73] and [74], researchers suggest that humans can be more effective when the environment and goals are in their best interest.

Other human performance attributes such as *psychomotor processing*, *spatial processing*, *composure*, and *perseverance* are important to improve the team cohesion through human performance enhancement. Overall personal (physiological, cognitive, and psychological) attributes have been classified into five subdomains after a detailed study by several defense agencies and are summarized in Table 2 [75]–[78].

2) MACHINE METRICS

All the machine-level metrics related to HMT, such as efficiency, performance, and accuracy, are well represented

TABLE 2. Human performance metrics for HMT.

Sub Domain	Attributes
Physical Health	general health; stamina; stress; fatigue
Cognitive	cognitive proficiency; attention; spatial processing;
Perception	memory; psychomotor processing; reasoning
Intra-Personal	composure; resilience; self-certainty;
	conscientiousness; success oriented; perseverance;
	decisiveness; impulsiveness; cohesiveness;
	assertiveness; adaptability; self-confidence
Inter-Personal	extraversion; judgment; team oriented; adaptability
Motive	moral interest; occupational interest

in literature. A few more are detailed as follows: *machine self-awareness*, or the degree to which a machine is aware of itself (limitations, capacities), is a precursor to reducing the human cognitive load and measured based on *autonomous operation time*, the *degree of autonomy*, and *task success* [62], [79]. Technically, *unscheduled manual* operation time may either be an interruption period in current plan execution or an unexpected assigned task [80]. *Neglect tolerance* (NT) is interpreted in numerous ways, such as machine performance falling below expectation, time to catch-up, the idle period, or operation time without user intervention. *State metric* helps track the machine or plan state based on four dynamic states: assigned, executed, idle, and out of the plan. *Robot attention demand* (RAD) is a measure of the fractional “task time” a human spends to interact with a machine. *Fan out* (FO) is a measure of how many robots with similar capabilities a user can interact with simultaneously and efficiently and is inverse of RAD [81]. *Interaction effort* (IE) is a measure of the time required to interact with the robot based on experimental values of NT and FO and is used to calculate RAD [81], [82]. Although humans can communicate through visual cues, gestures, etc., most machines need accurate information to act. Such information is mostly sent over wireless channels for various cyber-physical or cloud-robotics systems such as UAVs [83], [84]. Studies suggest that communication with machines in real time can be accomplished successfully by adapting to 5G communication technologies in hardware and software implementation [85]–[87].

Additional machine metrics that do not have a quantitative representation yet, and are difficult to measure, include *resource depletion*, *subgroup size*, *collision count*, *usability*, *adequacy*, *sensory-motor coefficient*, *level of autonomy discrepancies*, *physical constraints*, and *intellectual constraints* [20], [61], [80].

3) TEAM METRICS

Conventionally, a *team* has two primary components: a leader and one or more team members. A team leader is someone who provides guidance and instruction and leads the group to achieve set goals. In contrast, a team member is an individual who works under the supervision of a team leader [88]. Although there is no quantitative validation or representation of a team member, many guidelines and studies define the characteristics of an ideal team member who serves as a reference to evaluate or prepare a machine as a team member.

Five essential features of a human team member are defined as follows: functional expertise, teamwork, communication skills, job assignment flexibility, and personality traits [89]. In contrast, a well-defined or established machine team member feature list seems to have not been researched well due to the nascent nature of HMT research.

The key focus of team metrics is mission assignment and execution. *Task difficulty* represents the mental load a particular task generates [90]. The task difficulty metric for a machine depends on FO and requires three factors for measurement: *recognition accuracy*, *situation coverage*, and *critical time ratio* of a machine [65]. *Recognition accuracy* is the ability of the machine to sense its I/O parameters. *Situation coverage* (SC) is the percentage of situations encountered and accomplished by the robot. SC is defined based on *plan* and *act* stages of the mission. *Critical time ratio* is the ratio of time spent by a robot in a critical situation to the total time of interaction [65]. *Network efficiency* is the rate of flow of information between the human and the machine and determines the efficiency of interaction. It also influences time taken for scheduled and unscheduled manual operations, accuracy of mental computation, negligence tolerance, and human-machine ratio [20]. Four well-known subclasses of *false alarms* are true positive (TP), true negative (TN), false positive (FP), and false negative (FN) [63]. While false alarms measure complex communication between humans and machines in a team, people may ignore false alarms. A human factor study presented a trade-off between ignoring false alarms and *misses* and concluded that alarms are strongly situation dependent [91]. Some other team metrics that can be used in effective interactions are *hits*, *misses*, *automation bias*, and *misuse of automation* or metrics based on application scenario [92]. *Robustness* measures the ability of the team to adapt to the changes in task and environment during task execution [93] while *productivity* measures productive time compared to total invested time. *Task success ratio* indicates the number of completed versus allocated tasks [80].

Additional team metrics include *team effectiveness*, *human-robot ratio*, *cohesion*, *neighbor overlap*, *total coverage*, *critical hazard*, *autonomy discrepancies*, *TP*, *TN*, *FP*, and *FN interaction rates* (*TPIR*, *TNIR*, *FPIR*, *FNIR*), *cognitive interaction*, *cryptic coefficient* and *degree of monotonicity* [20], [94].

B. METRICS META-ANALYSIS

To identify common metrics, we need to analyze the metrics for properties such as aspect of measure, measurement technique, reliability and dependability of measurements, performance, and suitability for selected application area. These characteristics are identified through meta-analysis.¹ Metrics can be primarily classified based on either the measurement technique (subjective, objective, direct, indirect, nominal, ordinal, interval, ratio, process, resources, and results), or

¹ In this meta-analysis we study well-known published research works and reviews and identify the metric types defined here.

TABLE 3. Functional classification of HMT metrics.

Functionality	List of Metrics
Efficiency metrics	attention allocation; decision accuracy; mental workload; mental computation; workload; mental models; usability; sensory motor coefficient; plan execution; interaction efficiency; monotonicity; effort; cryptic coefficient; network efficiency; accuracy and coherence of mental models; recognition accuracy; fan out; span of control; flexibility; level of autonomy discrepancies; false alarms; true positive interaction rate false positive interaction rate; true negative interaction rate; false negative interaction rate; collision count; percentage request by operator; percentage request by machine; mode error; team productivity
Timing metrics	neglect tolerance; critical time ratio; autonomous operations time; manual operation time; scheduled operation time; unscheduled operation time; completion time; execution time; productive time; team performance; task success; intervention response time; intervention frequency; mutual delay; settling time; operator to robot time ratio; Mean Time Between Interventions (MTBI); Mean Time Completing an Intervention (MTCI); Mean Time Between Failures (MTBF)
Mission metrics	reliability; trust; total coverage; task allocation; plan state; plan execution; plan idle; plan out; neighbor overlap; similarity; task difficulty; situation coverage; robot attention demand; resource depletion and task success
Safety metrics	Risk to human; general health; critical hazard; fatigue; stress; self-awareness; human awareness; situation awareness

the quantity they measure (efficiency, safety, cognition, and time) [95], [96]. Here, we analyze the identified metrics based on measurement techniques, reliability, and performance and classify them as functional, subjective, objective, and real-time.

1) FUNCTIONAL CLASSIFICATION

Through this review, we found that several identified metrics can be employed in all three HMT agents with subtle modifications in measurement techniques; for example, the time taken by a human to complete a task can be measured using an external observer.² In contrast, machines use an automatic timer for the same purpose. We identified efficiency, time, mission, and safety as four functional classes of HMT metrics, as shown in Table 3. Metrics to evaluate efficiency will give the observer the required V&V to tune each agent to operate with maximum efficiency [20], [62], [63], [80], [81], [93], [97], [98]. Time metrics provide data related to the time taken for different operations by machine, human, and team, and these metrics are very important in *decision-making* and *performance* and *status determination* [20], [62], [65], [80], [81], [99]–[101]. Mission metrics measure attributes related to a task such as planning [20], [65], [80], [81]. Safety of the team is the highest priority for any mission, especially in stochastic and dynamic mission environments. Safety metrics measure the agent and mission safety during task execution [64], [71], [72]. Another class of metrics, termed as applied metrics, deals with the practicality and research on metrics and is divided into research and non-research metrics. Table 4 classifies the applied metrics with respect to the HMT agents.

2) SUBJECTIVE

Subjective metrics (SM) are used to measure abstract qualities based on human perception. These metrics may include feedback or judgment from observers (superiors or experienced professionals), for example, self-feedback, evaluation, or ratings.

Table 5 summarizes a few available well-documented SM scales. *Adaptability* is measured using a five-scale

²Observer is defined as a human or equipment with methods and tools to monitor the operation, performance, and progress of an HMT and provide standard feedback to improve HMT performance.

TABLE 5. Scales for subjective metrics.

Scales	Subjective metrics
Rathus assertiveness scale	Assertiveness
Connor-Davidson resilience scale; student motivation scale; resilience scale for adults	Resilience, Self-Confidence, Composure
Chernysenko scale	Conscientiousness
Big 5-factor model, Eysenck, HEXACO	Extraversion
Bratts Impulsiveness scale-11	Impulsiveness
Situation Awareness Global Assessment Technique (SAGAT)	Situational Awareness
Kuder occupational interest survey	Interest
Motivation-perseverance-grit scale	Perseverance
NASA Task Load Index	Workload

rating from the experts [102]. *Assertiveness* is measured based on the Rathus assertiveness scale [103], [104], while *resilience*, *composure*, and *self-confidence* are measured using 19 different scales, such as the Connor-Davidson resilience scale, student motivation scale, and resilience scale for adults [105], [106]. *Conscientiousness* is computed using the Chernysenko scale, which is a 60-item question inventory, with each question rated by subjects on a 4-point scale [107]. *Decisiveness* is measured with subject ratings on the need of information, confidence in decision-making, and self-appraisal. It is also notable that peers can rate subjects' decisiveness as well [108].

Extraversion is measured using various rating scales such as the Big Five-Factor Model, Eysenck, and HEXACO [109], [110]. The *emotional state* of a person is calculated by ratings on behavior, facial expressions, and startle response [111]. *Impulsiveness* is measured using Bratts Impulsiveness scale-11 (BIS-11), which is the 11th version of the original 30-question inventory proposed by Bratt in 1985 [112]. *Situational awareness* is measured using the simulation technique called Situation Awareness Global Assessment Technique, which includes subjective inputs as well as objective measures [113]. *Perseverance* is measured by the scores obtained from the motivation-perseverance-grit scale that requires self-ratings [114]. *Human awareness* can be measured on a scale with the help of self or expert ratings [62], [115]. The workload is calculated using a multidimensional self-rating scale, for example,

TABLE 4. Applied metrics.

Human	Machine	Team
Adaptability	Usability	Cohesion
Assertiveness	Fan Out	Interventions
Impulsiveness	Robot Attention Demand	Intervention Response Time
Cohesiveness	Collision Count	Neglect Tolerance
Perseverance	Plan Execution	Unscheduled Operations Time
Extraversion	Plan Idle	Time Autonomous Operations
Conscientiousness	Plan out	Time in Manual Operations
Humility	Plan State	Plan State
Occupational Interest	Resource Depletion	Situation Coverage
Psychomotor processing	Interaction Effort	Task success
Stamina	Mutual Delay Time	Task Difficulty
General health	Neglect Tolerance	False Alarms
Fatigue	Settling Time	False Positive Interaction Rate
Stress	Time in Autonomous Operations	False Negative Interaction Rate
Situation Awareness	Time in Manual Operations	Interaction Efficiency
Attention Allocation Efficiency	Unscheduled Operations Time	Network Efficiency
		Recognition Accuracy
		Team Productivity
		True Negative Interaction Rate
		True Positive Interaction Rate

□ Research Metrics ■ Non-Research Metrics

the NASA-TLX [62], [116]. Among machine metrics, self-awareness and adequacy are SM, as they require human expert ratings on deviations [61].

Table 6 illustrates the pros and cons of a few popular self-reporting scales. One of the biggest drawbacks of SM is being biased in self-reported scales. For example, individuals with high neuroticism traits are expected to report more distress, pressure, etc., than others [117]. Other biases may include different socioeconomic strata, introspective ability, and image management [118], [119].

3) OBJECTIVE

Objective metrics (OM) are task-specific tools, functions, and formulae to measure task performance quantitatively. OM are developed to measure an activity that can be changed, customized, or expressed by a value for comparison [120]. Most identified machine and team, as well as a few human, metrics, are OM. In human metrics, *general health* can be considered an objective measure because it is measured by recording blood pressure, temperature, and heart rate [61], [121]. Similarly, *physiological fatigue* can be measured using heart rate, blood pressure, galvanic skin response, and adrenaline level. *Visual fatigue* is calibrated using Swedish occupational fatigue inventory, which employs parameters such as cardiovascular response, energy expenditure, skin temperature, and blink rate [122]. *Stress* is measured as a function of blood pressure, vocal tone, salivary alpha-amylase levels, heart rate, and blood cortisol levels [123]. *Stamina* measurement may involve taking into account parameters such physical activity (push-ups and running-speed [61]), *shift length* (the time span in which one needs to be attentive [124], [125]), or *vigilance* (through traditional human factors or modern eye-tracking methods [126]).

The *memory* of an individual is measured by the degree of *recognition*, *relearning*, and *reconstruction* that is determined using the formulae to measure memory [127]. *Cognitive proficiency* is measured using the cognitive proficiency index, which is defined as an auxiliary scale by Wechsler intelligence scales [128].

Various *time metrics* such as *intervention response time* (time taken by the human to intervene if a problem occurs) [20], *overhead time* (time spent by the machine in idle state or unplanned activities) [80], and *productive time* (cumulative sum of time spent by the team in scheduled manual, unscheduled manual, and autonomous operations time) are also relate to objective metrics. *Neglect impart* (NI) is calculated from the NT graph by measuring the *neglect time*, or the average time before the robot's performance falls below a threshold [68]. *Settling time* is the time taken to reach the required accuracy by the machine [100]. In contrast, *completion time* is calculated for the time taken by an HMT to complete a given task. The *critical time ratio* is the ratio of the duration of the critical mission section to the duration of interaction [65]. *Task success* is calculated as the percentage of the successful tasks [80].

The *decision-analysis* approach follows a Bayesian view of probabilities associated with the possible events, making it an objective measure [62], [129]. *Inferred mental workload* takes into account eye movement activity, cardiac functions (ECG), brain activity (EEG), and Galvanic skin response (GSR) [61]. Previously discussed metrics such as attention allocation, situation coverage, state metrics, false-alarm metrics (TP, TN, FP, and FN), RAD, and IE are also can be objective. *Human trust and reliability* on a machine are derived (i.e., inferred) from its FO factor and RAD. As the RAD increases, the user trust and reliability on the machine

TABLE 6. Pros and cons of using subjective scales.

Scale	Pros	Cons
Connor-Davidson resilience scale (CD-RISC); and student motivation scale;	The scale is well defined, and the factor analysis of this scale yielded the big five factors. The scale also demonstrates that with proper training resilience can be improved [105, 106].	The scales focus on resilient qualities at the individual level and these scales prompt speculation.
Chernyshenko scale	Uses unified scores of 6 major factors, with each factor scored using analysis of 7 personality inventories, for conscientiousness computation. In-depth analysis of its effect on human performance studied in [107].	Difficulties in assessing facets and measuring through scales due to their non-orthogonal nature.
Big Five-Factor model; the smaller seven; HEXACO	These models define the personality traits of a human, which have been used in designing scales for human performance as an SM.	These scales prompt the user to speculate in self-reporting.
Bratts Impulsiveness scale-11	The score obtained can be used to calculate impulsiveness, which can in turn help in assessing the human performance [112].	Self-reporting limitations that leaves room for speculations.
Situation Assessment Global Assessment Technique (SAGAT)	SAGAT is a well-documented tool to measure an SA, possesses a high degree of content validity based on the SA requirements analyses and is used to create the queries that were found to have predictive validity.	Limited to simulation environment most of the times.
Motivation-perseverance-grit scale	Grit scale enables prediction of perseverance and motivation for long-term goals. It was found to be the best predictor among many other indicators of which cadets will drop out after first difficult summer training [114].	Self-reporting limitations that leaves room for speculations.
The NASA Task Load Index (NASA-TLX)	Initially introduced in 1984, efforts have been put in to make it more flexible and robust. Use of this metric since its inception is well-studied [62, 116].	Self-reporting limitations that leaves room for speculations.

TABLE 7. Parameters to metric mapping.

Parameters	blood pressure	body temperature	heart rate	galvanic skin response	adrenaline production	cortisol levels	salivary alpha-amylase	vagal tone	EEG	ECG	eye movement and blink rate	energy expenditure	overhead time	intervention response time	autonomous operations time	manual operations time	unscheduled operations time	neglect time	Settling time	time to complete	interaction frequency	true positive	false Positive	true negative	false negative	interaction time	plans level
Metric																											
General health																											
Physiological fatigue																											
Visual fatigue																											
Stress																											
Productivity time																											
Neglect Impart																											
Attention allocation																											
Mental workload																											
Interaction effort																											
TPIR																											
FPIR																											
FNIR																											
TNIR																											
RAD/FO																											
State																											

decreases, for example, IE, and NT inversely affect human trust and reliability. Another OM, *total coverage*, is a measure of the area or environment used by all the sensors simultaneously at a specific time during the mission execution [94]. *Neighbor overlap* can help measure how much a machine affects the performance of other machines. *Network efficiency* can be measured using bandwidth and latency.

All identified objective metrics are presented in Table 7, mapping the metrics to their corresponding parameters.

Researchers can use this table to identify redundant parameters and eliminate bias.

4) REAL-TIME METRICS

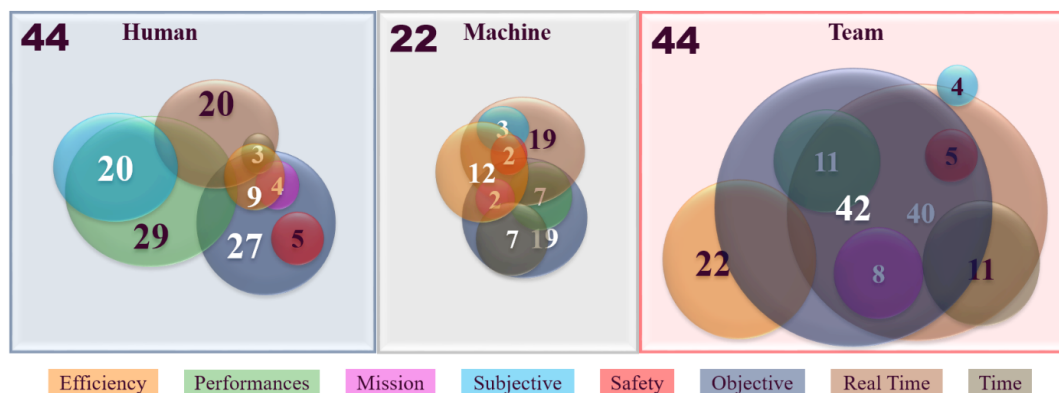
Real-time metrics are crucial in any time-sensitive, real-time applications such as engineering, defense, and healthcare. Purposes include, but are not limited to, improving communication, response times, information transfer accuracy, and mission success rate.

TABLE 8. Comprehensive color-coded classification matrix of HMT metrics.

Human		Machine	Team	
Adaptability	Success oriented	Self Awareness	Cohesion	Situation Coverage
Assertiveness	Psychomotor Processing	Adequacy	Cognitive Interaction	Task Allocation
Composure	Situation Awareness	Usability	Coordination Demand	Task Difficulty
Conscientiousness	Stress	Decision Accuracy	Elongation	Task Success
Decisiveness	General Health	Fan Out	Human Robot Ratio	Total Coverage
Extraversion	Human awareness	Sensory motor Coefficient	Robot Attention Demand	Similarity
Emotional State	Fatigue	Collision count	Critical Hazard	Autonomy Discrepancies
Impulsiveness	Intervention Frequency	Plan State	Robot Awareness	Monotonicity
Moral Interest	Productive Time	Network efficiency	Human Risk	Effort
Occupational Interest	Overhead Time	False Alarms	Critical Time Ratio	False Alarms
Perseverance	Reliability	Robot Adaptability	Interventions	FPIR
Reasoning	Trust	Resource Depletion	Interventions Time	FNIR
Resilience	Total Coverage	Interaction effort	Neglect Tolerance	Flexibility/Robustness
Self-Confidence	Attention Allocation	Mutual Delay Time	Unscheduled time	Interaction Efficiency
Self-Certainty	Decision Accuracy	Neglect Tolerance	Productive time	Network Efficiency
Cognitive Proficiency	Mode Error	Settling Time	Time Autonomous Operations	Recognition Accuracy
Cohesiveness	Span of Control	Time Autonomous Operations	Time in Manual Operations	Requests by Operator
Judgment	Mental workload	Time in Manual Operations	Time to complete	Requests by Robot
Memory	Mental Computation	Unscheduled Operations Time	MTBI	Team Productivity
Spatial Processing	Workload	RAD	MTCI	TNIR
Stamina	Guesses to Success	Errors	Neighbor Overlap	TPIR
Team oriented	Mental Models	Navigation	Plan State	Cryptic Coefficient

Performance
Safety
Efficiency
Subjective
Objective
Mission
Time

Real Time
Contact
Invasive

**FIGURE 4.** Quantitative graphical representation of overall metrics classification (summary of Table 8).

Psychomotor processing calibrates human psychomotor speed during the mission along with *spatial processing*. These, along with stress, fatigue, general health, and various time metrics, are also known as real-time metrics. Although situation and self-awareness are crucial for a system, they cannot yet be measured in real time. Subjective measures and human attributes or traits are tough to measure in real time because they depend on the observer scale measurement, for example, adaptability, assertiveness, composure, cognitive proficiency, conscientiousness, and decisiveness. Few other metrics such as memory, decision accuracy, autonomy discrepancies, and cognitive interaction can be evaluated only

after the accomplishment of the mission. Research reviewed did not indicate any of these being measured in real time even though there is a possibility of real-time measurement through recent developments in prediction models and computing. Therefore, these can be classified as non-real-time metrics [130].

5) SUMMARY

In general, subjective and objective measurement techniques do not measure the same parameter. However, there are a few parameters, such as cognitive load and stress, which may employ both of these measurement techniques. Based

on the accuracy of the technique, one might be preferred over the other. For example, subjective measurement techniques work better on task load despite the availability of objective measurement techniques [106], [107]. Subjective metrics are recommended in combination with objective metrics for human performance. At the same time, we avoided metrics which are either derivatives of, or involve parameters similar to other metric(s). Thus, ensuring that the selected common metric set will somewhat represent those avoided metrics during the HMT performance evaluation.

Table 8 summarizes this section as a color-coded matrix representing a taxonomy in which one can look for popular metrics, relationships among metrics, selection of metric relevant to HMT agent, measurement techniques, and measurement aspects. Such a taxonomy is expected to allow the research community to study HMT metrics and develop a better set of common metrics. The metrics in bold are the common metrics we identify and discuss in detail in the

next section. Figure 4 summarizes the color-coded table quantitatively and shows the total number of metrics represented in the table, the different aspects they measure, and measurement methods.

IV. COMMON METRICS FOR HMT BENCHMARKING

It is understood that establishing a set of common metrics for all possible types of HMT is difficult and may not enable benchmarking for every application. Keeping that in mind, we define a set of metrics that are common to selected application areas. Nonetheless, this set may apply to a wider range of tasks or areas. Although several works attempted to identify HMT applications, our survey found only a few to either establish common metrics or at least provide guidelines for such an identification [62], [131]. Customary practices include identifying common metrics from experience, using metrics that researchers are familiar with, or attempting to measure all available aspects of a system. These approaches may lead to inefficiency due to the possible use of inappropriate measurement methods, cost of implementation, or lack of strong face validation of a measure.

In [132], researchers proposed a set of common metrics to measure the performance of interaction with the limitation of targeting only the robot or the human. Researchers identified common metrics for three agents using subjective rating scales for HRI [62], [133], which come with their share of limitations such as less performance estimation accuracy, poor reliability, spillover effects, and perspective measures (that vary based on perspective) [134]–[136], and which can sabotage the entire benchmarking. Another earlier work detailed a supervisory control system and proposed generalized metrics for specific examples, such as single human and HRI for multi-robot teams [137]. Later, a set of metrics for measuring supervisory control performance was selected [96]. Selection criteria of proposed common metrics are listed in Table 9. To summarize, each metric is selected based on five major aspects: total attributes a metric represents, measurement method, strong face validation of

the metric, well documented in literature and practice, and supports the selected applications. Moreover, metrics must represent a team dynamics rather than an individual agent.

A. COMMON HUMAN METRICS

The common metrics for human performance should give an analytical representation of human performance in an HMT. For common human metrics, we eliminated all the metrics that are invasive and only subjective to make the measurement practical. In additions to the selection criteria defined above, we focused on measurement methods for human metrics because relating activity measures to human performance is difficult [75]–[78]. Our research also agrees with that of several researchers in presenting trust, cognitive load, and human fatigue as important HMT metrics. However, due to lack of concrete objective measurement methods, and a strong correlation between resulting measurements and HMT performance, we excluded those from our selection. We identified four potential common human metrics: judgment, attention allocation, mental computation, and mode error.

1) JUDGMENT

Judgment, or decision-making, is the process of observing and assessing situations, drawing conclusions, and predicting action consequences. It can be measured subjectively, objectively, or via mixed measurement methods. In HMT, judgment can be classified as situational or practical and may require measurement while selecting a human teammate, and in team-building and task execution, respectively. Using a combination of measurement techniques can yield a better result, including up to 90% accuracy in measuring judgment [138]. Compared to practical judgment, multiple studies have been carried in fields such as healthcare and defense to measure situational judgment. With test samples ranging from 1200 to 10600, most tests yielded accurate results [138]–[144]. Limitation of the method includes simulations not being representative of a practical scenario. In addition, it is known that an individual may compromise judgment for an experimental scenario and judge differently in the real world [138], [139]. Further, the *Test of Practical Judgment* [145] was found to be a prominent test for safety, social and ethical issues, and financial issues, with 134 samples showing promising results. However, the existence of only a few studies that used this method indicates a lack of widespread usage. Judgment is a mission metric that directly correlates to the human performance and efficiency in an HMT. As described above, judgment is well researched and has various studies proving the correlation with reliable results. Moreover, judgment as a mission metric represents human action in an HMT and should be able to provide the human factor analysis needed in HMT benchmarking.

2) ATTENTION ALLOCATION

In stressful situations with complex systems, it is possible that focus is shifted from an important task to a minor or an unimportant task [69]. Therefore, it is expected that

TABLE 9. Common metric selection criteria.

Items analyzed	Attributes	Selection criteria
Category	Safety, efficiency, time, mission and performance	Performance and efficiency (preferred if it includes others)
Measurement method	Subjective and/or objective, invasive and non-invasive	At least one objective method and must be non-invasive
Research performed	Publish research report, reviews, project data	15+ peer reviewed publications considered
V&V of results	Sample size in user evaluation, accuracy of measurement, agreement in results	High sample size is preferred but multiple evaluation is mandatory
Application scenarios	Search, navigation, target identification, ordinance disposal, geology, surveillance, healthcare training, and tour guiding	Must be applicable in all application scenarios, measuring method can vary

tracking real-time attention allocation will improve an HMT. A 2008 review discussed a few attention metrics including eye tracking, verbal protocols, and tracking resource allocation cognitive strategies (TRACS) [61]. Several studies were performed in TRACS, with a maximum participant size of up to 45, showing a correlation with attention allocation. TRACS is achieved by measuring HCI with a 2D representation of a human [146], [147], and a common limitation involves customization for each interface and task [61]. Various researchers have studied and correlated eye tracking, attention allocation, and human performance using fixations, saccades, pupillometry, and blinks for application areas such as UAVs, supervisory control, and healthcare [61], [148]–[151], while encountering limitations such as limited correlation between gaze and thinking, intensive data analysis, and noise in the measured data [152]. In conclusion, an effective measure of attention can be achieved through combining eye tracking and TRACS. As described above, attention allocation is a well-studied metric that has different measurement methods and satisfied the criteria to be selected as a common metric. In addition, it is noteworthy that attention allocation deals with human parameters that directly affect HMT performance.

3) MENTAL COMPUTATION

Mental computation, mental workload, and cognitive load are well-studied theories and recent studies establish their correlation with human performance [153], satisfying our criteria of common metric selection. However, since mental computation is a non-real-time metric, HMT developers need to perform mental computational studies and adjust their design for peak performance. Primary measurement methods use subjective scaling and physiological performance parameters. Performance measures can be used to measure relative speed, accuracy, and elapsed time [153]. Physiology studies involving mental computation and human performance are overwhelming, as studies include EEG, ECG, GSR, eye tracking, etc., with participants ranging from 28 to 300 and tasks ranging from defense and medicine to controls [154]–[159]. Studies successfully differentiated between multiple and increased mental workload based on task demand but failed to show a consistent correlation between efficiency and identified cognitive load patterns. Therefore, these measurement methods should be used only

for minimizing mental workload at the HMT design stage, which may result in better performance during operation.

4) HUMAN ERROR

An error has a direct correlation with performance and efficiency. Every system needs to rectify mistakes if any in real-time and post-mission completion. Along with the previously mentioned selection criteria, the above two factors led to the selection of human error as a common metric. This category of error is one of the most prominent and important metrics. There are several types of errors presented in the literature. However, *mode error* is one of the most studied metrics. Mode error represents the human error that affects HMT operation. Mode error is defined as the difference in actual and intended operation mode as a result of either a human-machine miscommunication or a human selecting an incorrect mode of operation [99]. Mode error is a widely studied and prominent human error that can affect the human-machine relationship and depends on the application scenario. Mode error can adversely affect performance based on the severity of the error. If unchecked, a mode error may result in total system failure. Researchers have measured mode error during system operation in various ways [99], [160], [161]; for example, mode of operation must change when flying conditions change while flying a single-engine airplane with focus on airspeed, altitude, and routing by controlling thrust, ailerons, elevators, and rudder. Otherwise, a mode error occurs and might leads to catastrophic system failure. Mode error can be converted to an empirical value for some applications. However, it is noteworthy that all scenarios cannot be easily generalized.

B. COMMON MACHINE METRICS

To select a common metric from the identified metrics, an application-specific primary analysis was conducted. A machine parameter can be easily measured; however, identifying a metric that may apply in broad application space is quite challenging. In this section, we identified metrics that have a maximum number of mutually exclusive parameters in addition to the selection criteria defined in Table 9. The goal is to provide metrics that measure performance, efficiency, and accuracy of task operations while minimizing parameter redundancy. We have identified three potential common machine metrics described below.

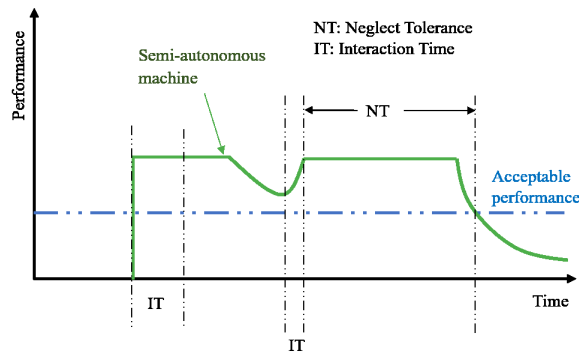


FIGURE 5. Neglect tolerance (NT) model.

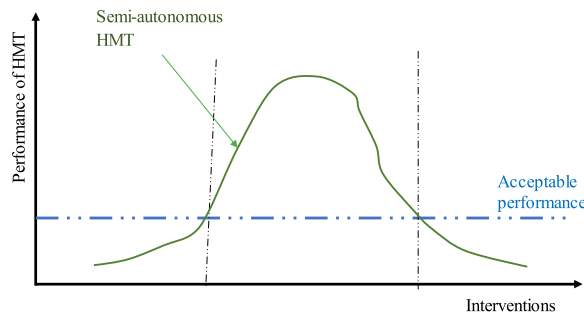


FIGURE 6. Hypothetical interventions curve.

1) ROBOT ATTENTION DEMAND (RAD)

RAD represents the relationship of the machine with human teammate and is measured using NT and IE, as shown in equation 1 [132], [162], [163]. We further discuss these in detail.

$$RAD = \frac{IE}{IE + NT} \quad (1)$$

- **Neglect tolerance (NT)** is a unique characteristic graph of machine performance that is measured for each autonomous system individually, as shown in figure 5 [132], [164]. NT usually decreasing trend with time, while the rate of change varies from machine to machine [132], [162]. Although no standards have been established or adopted for NT measurement, several researchers have adopted NT in their studies [122], [153], [155].
- **Interaction effort (IE)** is the capability of the machine to understand the higher human communication level. It is not just a physical input or account for stages of understanding information and decision-making; it can be inferred from secondary parameters. For example, eye tracking can be used to determine whether a human was looking at the display before an input. Therefore, including time for these tasks would be more accurate. A hypothetical interaction effort characteristic graph is one of the prominent models many researchers have adapted, where IE is estimated using RAD and interaction time [162]. However, in practice, researchers have measured the interaction time as the IE [163].

In relating RAD to autonomous systems' performance, studies state [132] and experimentally found that a lower RAD value results in a better performance [162], [163]. Most of the experiments conducted in this area are related to robots and software agents. The metric has not been evaluated with humans or an HMT. RAD is another well-documented metric in HMT literature that can measure machine performance and real-time efficiency while the machine is operating in an HMT setting. RAD is a unique metric developed solely for machine teammates in an HMT. Even though primary results satisfy the selection criteria for RAD, several mathematical models suggest further possible development.

2) MACHINE STATE METRIC

Machine state metric was coined and first used for an airplane in 2010 [80]. However, measuring or identifying a machine state and its changes is a widespread practice. Possible states for a given autonomous system are represented as a state chart, and popular types include the rendezvous manager state chart (four states), data flow manager state chart (five states), and unified modeling language statechart (varying number of states) [165]–[167]. State measurement is helpful in real-time system observation and correlating machine clock time and machine performance. Machine state can also provide a sense of a machine's operation level and facilitate monitoring. Following is an example of how the state of a machine could be measured. For example, if a machine has four states—assigned, executed, idle, and out of plan—only one state can be true at a single point in time, and enumerated task status values are indicated as the following: failure = >0, successful = 0, executing = -1, paused = -2, and pending = -3 [80], then:

$$\text{State at time } t = \begin{cases} \text{plan assign status} = -3 \\ \text{plan execution status} = -1 \\ \text{plan idle state} = -2 \\ \text{out of plan status} \geq 0 \end{cases}$$

3) ERRORS

The error is one of the prominent metrics that can be represented in several ways based on the machine and application type. An error has a direct correlation with machine performance, efficiency, and task success. Machine errors also affect the team performance in several ways, ranging from affecting user trust to increasing workload, stress, and fatigue. If the errors are high and frequent, creating an HMT would be counterproductive to the mission or task. Machine errors include various types of faults and defects and vary based on the application [168]. For example, hardware fault in a vehicle may lead to a hardware error, while *interpretation error* appears due to the environmental conditions, which are difficult to model [169]. Researchers have also described a few other errors related to software intelligent assistants such as interaction errors, data entry errors, cumulative calculation errors, cognitive overload, misrepresentation of information, and security errors [170], [171]. Error correction

methods rely on error types, and the popular methods include simulation, modeling, testing, verification, and validation; for example, modeling errors can be avoided using simulation while hardware faults are identified through verification during design and development [169], [172], [173]. There is no empirical formulation for errors, in general; however, based on application and error type, a unique value can be awarded to an error to represent the effect of the error on performance.

C. COMMON TEAM METRICS

After careful analysis, we identified three potential common team metrics that can be used in measuring the performance of a team or system that satisfies the selection criteria established in Table 9. The combination of these three metrics, along with human and machine metrics, will provide an overall HMT performance score level and are discussed below:

1) PRODUCTIVE TIME

Measuring time is a relatively simpler and more reliable way to achieve higher accuracy when compared with other metric measurement techniques. Productive time is a metric that is used widely to measure team productivity by measuring the time spent by the machine and the human on a mission, and it is represented by the following equation:

$$\begin{aligned} \text{Productive time} = & \sum \text{Autonomous operation time} \\ & + \text{Manual operation time} \\ & + \text{Unscheduled manual operation time} \end{aligned}$$

Many researchers have used productive time and total task time to measure their robot and teams' performance [8], [80], [174], [175]. For example, in a task of transferring an object from location A to location B, productive time involves object retrieval time, travel time, and replacement time. Other times such as planning, rerouting, and delays in communication will be added to task completion time but not productive time. Productivity is calculated as the ratio of productive time and total task time. Productive time evaluates team productivity and efficiency, which is a key parameter defining the team success. It is also a well-studied metric, measurable in real time, and can be objective with stronger face value. All of these contribute to its selection as a common metric.

2) COHESION

Cohesion is defined as a dynamic process that is reflected in the tendency of a group to remain united in the pursuit of its instrumental objectives and for the satisfaction of member affective needs [176], [177]. Our survey identified hundreds of published research studies and books on cohesion in human teams, indicating its importance in team performance evaluation. We have selected it as a common metric based on our review of both human team studies and HMT studies because of cohesion's strong and direct relation

to team performance. It also satisfies all the requirements established in Table 9 and represents the effect of a team on each HMT agent. Cohesion has been studied in human teams to improve team performance since 1978. The group environment questionnaire is a widely used self-reporting subjective method for measuring cohesion [177]. According to a review, cohesion demonstrated a significant effect on team performance [178] and can be measured unidimensionally or multidimensionally, with the latter being better. As an SM, it is difficult to incorporate it in real time. However, team measurements prove that cohesion is a function of time and plays a key role in measuring the extent to which a team can work together before deploying a team. A standard method of measuring cohesion in HMTs has still to be developed. Several researchers used communication patterns between teammates and connected members to measure cohesion [94], [179], [180]. In conclusion, cohesion plays a key role as a metric to measure team performance or teaming nature.

3) INTERVENTIONS

Although human intervention may have a negative impact on the overall HMT performance, it is necessary to resolve errors made by a machine. It has been used widely to represent autonomous system performance and correlates the number of interventions to the performance of a machine [181]. Timely and optimal number of interventions by a teammate will lead to better performance in an HMT [182]. It can be measured in intervention time or intervention response time, which is measured as the total time a machine spent responding to interventions. In both methods, simple timers or counters can be used [62], [183]. However, this metric needs more in-depth studies before a standard to measure HMT performance can be developed. Based on analyzed studies, we hypothesize that intervention is a nonlinear function with an inverted U-shaped dose-effect curve drawn against a performance of an HMT as shown in figure 6. For example, too many interactions in a synthetic assistant-based learning environment may cause interruptions in learning while too few may give reasons to repeat earlier mistakes.

D. DISCUSSION

Selected common metrics in all three agents of HMT may be helpful to measure the HMT performance and derive an empirical value to allow comparison with another HMT [184]. This measurement is multidimensional (application, scenarios, agent, etc.) and will give an in-depth analysis of difficulties in HMT applications that need to be improved to achieve better performance. We have identified 10 common metrics among more than 100. These metrics have many parameters as sub-metrics that allowed detailed HMT analysis for aspects such as safety, performance, and efficiency. In [131], researchers attempted to analyze the possibility of a machine as a teammate. Their concluding remark suggests that future HMT researchers may need to identify the uniqueness of a machine and design an HMT such

TABLE 10. Summary of identified common metrics.

<i>Metric</i>	<i>Agent</i>	<i>Selection criteria</i>	<i>Description</i>
Judgement	Human	Objective, non-real-time, user studied, reviews, and correlation with team performance	Measures human judgement skills and trust levels, can be measured at design or application stage
Attention allocation	Human	Objective, real-time, user studies, reviews, and correlation with human performance	Proven measurement techniques will measure human attention allocation efficiency that can be related to human performance
Mental computation	Human	Objective and/or subjective, user studies, reviews and correlation with HMT design	Using EEG techniques, we can create human mental models that will be useful in HMT design and development
Human Error	Human	Objective, real-time, published results, relation with task execution	Real-time mode error measurement will help HMT execute its tasks with precision
RAD	Machine	Objective, real-time, published results, characteristic graphs and relation with machines performance	RAD monitoring will provide significant results that can be used to measure machine performance
State	Machine	Objective, real-time, published results, characteristic graphs and relation with machines performance	Machine state can be used by human to understand the machine, and help improve machine and team performance
Errors	Machine	Objective, real-time, published results, characteristic graphs and relation with machines performance	Being a generalized metric that gives all machine errors as a quantitative value and can be used in performance evaluation
Productive time	Team	Objective, real-time, published results, characteristic graphs and relation with team performance	Being a time metric, it can be used to significantly identify team success
Cohesion	Team	Objective, real-time, published results, characteristic graphs and relation with team performance	An observer metric and helps in identifying HMT teaming nature quantitatively
Interventions	Team	Objective, real-time, published results, characteristic graphs and relation with team performance	Can be positive or negative in performance score formula and plays a crucial role in understanding team

that members (human/robot) complement each other rather than designing a system in which a robot merely imitates a human. In this context, selected common metrics may help measure each agent individually, measure the HMT independently, and help future engineers design tailored HMTs and benchmarks. Through this study, we found that the performance of the HMT is rated based on a performance-score, which is a weighted combination of common metrics [96], 184, [185]. This score can act as an application-specific HMT benchmark and provide a relative performance score, thus providing a platform for HMT comparison [186], [187].

V. CONCLUSION

Synthetic teammates are moving from personal voice assistants that answer questions such as “How’s the weather outside?” and set meeting reminders to assistants that can be used in healthcare, large-scale industrial production systems, military surveillance, threat neutralization, and national security. These application areas typically entail a threat to the human life along with huge investments that make use of a standardized HMT essential for task execution. To conclude, we would like to point out the importance of an application-specific HMT performance evaluation that could use the identified common metrics.

Through this review, we proposed a definition and identified the components and functional blocks of an HMT. At the beginning of the review, we posed three goals to achieve through this review: (i) identify available metrics in HMT, (ii) analyze and classify identified metrics, and (iii) propose common metrics. Available metrics were identified in section 3.1; analysis and classification were achieved in section 3.2; and finally, we proposed 10 common metrics to evaluate HMTs in section 4. The common metrics have also been summarized in Table 10. Metric versus parameter table, and color-coded metrics table are ancillary results of the review.

Although a common metric may be used for various applications, the interpretation of the scores might be application

specific; for example, a UAV HMT will have a different scoring mechanism than a healthcare assistant. In conclusion, selecting appropriate test populations when benchmarking an HMT is very important. Specifically, as robots are increasingly deployed in applications where the target user is not an expert roboticist [188], it becomes critical to recruit subjects with a broad range of knowledge, experience, and expertise. The continuing work under this effort will expand and refine the material presented here. The eventual plan is to provide a living, comprehensive document that future research and development efforts can utilize as an HMT metric toolkit and reference source.

ACKNOWLEDGMENT

The team would like to thank Colin Elkin, Ruthwik Junuthula, Naga Babu Koganti, and Mayukha Cheekati for their help in proofreading and editing.

REFERENCES

- [1] E. Schechter, *AUVSI Reporter: AFRL Targets Seamless Human-Machine Interaction*, V. Muradian, Ed. Vienna, VA, USA: Sightline Media Group, 2014. [Online]. Available: <http://defensenews.com>
- [2] B. C. Hoekenga, “Mind over machine: What Deep Blue taught us about chess, artificial intelligence, and the human spirit,” M.S. thesis, Dept. Humanities, Graduate Program Sci. Writing, Massachusetts Inst. Technol., Cambridge, MA, USA, 2007.
- [3] P. Griffiths and R. B. Gillespie, “Shared control between human and machine: Haptic display of automation during manual control of vehicle heading,” in *Proc. 12th Int. Symp. Haptic Interfaces Virtual Environ. Teleoper. Syst. (HAPTICS)*, Mar. 2004, pp. 358–366.
- [4] M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. M. Jonker, M. B. Van Riemsdijk, and M. Sierhuis, “Coactive design: Designing support for interdependence in joint activity,” *J. Human-Robot Interact.*, vol. 3, no. 1, pp. 43–69, 2014.
- [5] R. R. Hoffman, J. K. Hawley, and J. M. Bradshaw, “Myths of automation, part 2: Some very human consequences,” *IEEE Intell. Syst.*, vol. 29, no. 2, pp. 82–85, Mar./Apr. 2014.
- [6] M. Sullivan et al., *F-35 Joint Strike Fighter: Assessment Needed to Address Affordability Challenges*, DTIC document GAO-15-364, Apr. 2015.

- [7] J. S. Hicks and D. B. Durbin, *An Investigation of Multiple Unmanned Aircraft Systems Control From the Cockpit of an AH-64 Apache Helicopter*, DTIC document ARL-TR-7151, Dec. 2014.
- [8] J. Ball et al., "The synthetic teammate project," *Comput. Math. Org. Theory*, vol. 16, no. 3, pp. 271–299, 2010.
- [9] P. C. Cacciabue, "Elements of human-machine systems," in *Guide to Applying Human Factors Methods*. London, U.K.: Springer, 2004, pp. 9–47.
- [10] J. L. Burke, R. R. Murphy, E. Rogers, V. J. Lumelsky, and J. Scholtz, "Final report for the DARPA/NSF interdisciplinary study on human-robot interaction," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 2, pp. 103–112, May 2004.
- [11] T. J. Wiltshire and S. M. Fiore, "Social cognitive and affective neuroscience in human-machine systems: A roadmap for improving training, human-robot interaction, and team performance," in *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 6, pp. 779–787, Dec. 2014.
- [12] S. Nikolaidis, R. Ramakrishnan, K. Gu, and J. Shah, "Efficient model learning from joint-action demonstrations for human-robot collaborative tasks," in *Proc. 10th Annu. ACM/IEEE Int. Conf. Human-Robot Interact.*, Mar. 2015, pp. 189–196.
- [13] M. C. Gombolay, R. A. Gutierrez, S. G. Clarke, G. F. Sturla, and J. A. Shah, "Decision-making authority, team efficiency and human worker satisfaction in mixed human-robot teams," in *Autonomous Robots*, vol. 39, no. 3, New York, NY, USA: Springer, Oct. 2015, pp. 293–312, doi: [10.1007/s10514-015-9457-9](https://doi.org/10.1007/s10514-015-9457-9).
- [14] M. C. Gombolay, R. J. Wilcox, and J. A. Shah, "Fast scheduling of robot teams performing tasks with temporospatial constraints," *IEEE Trans. Robot.*, vol. 34, no. 1, pp. 220–239, Feb. 2018.
- [15] J. Kim and J. A. Shah, "Improving team's consistency of understanding in meetings," *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 5, pp. 625–637, Oct. 2016.
- [16] S. Nikolaidis, P. Lasota, R. Ramakrishnan, and J. Shah, "Improved human-robot team performance through cross-training, an approach inspired by human team training practices," *Int. J. Robot. Res.*, vol. 34, no. 14, pp. 1711–1730, 2015.
- [17] D. J. Bruemmer and M. C. Walton, *Collaborative Tools for Mixed Teams of Humans and Robots*, DTIC document ADA478926, 2003.
- [18] C. E. Harriott and J. A. Adams, "Modeling human performance for human-robot systems," *Rev. Human Factors Ergonomics*, vol. 9, no. 1, pp. 94–130, 2013.
- [19] M. D. Manning, C. E. Harriott, S. T. Hayes, J. A. Adams, and A. E. Seiffert, "Heuristic evaluation of swarm metrics' effectiveness," in *Proc. 10th Annu. ACM/IEEE Int. Conf. Human-Robot Interact. Extended Abstr.*, Portland, OR, USA, Mar. 2015, pp. 17–18.
- [20] C. E. Harriott, A. E. Seiffert, S. T. Hayes, and J. A. Adams, "Biologically-inspired human-swarm interaction metrics," in *Proc. Human Factors Ergonomics Soc. Annu. Meeting*, 2014, vol. 58, no. 1, Newbury Park, CA, USA: Sage, pp. 1471–1475.
- [21] S. D. Sen and J. A. Adams, "Real-time optimal selection of multirobot coalition formation algorithms using conceptual clustering," in *Proc. Workshops 29th AAAI Conf. Artif. Intell.*, 2015, pp. 1–7.
- [22] C. Lizza and C. Friedlander, "The Pilot's Associate: A forum for the integration of knowledge-based systems and avionics," in *Proc. IEEE Nat. Aerosp. Electron. Conf. (NAECON)*, May 1988, pp. 1252–1258.
- [23] G. Champigneux, "In flight mission planning in the 'copilote électronique,'" in *AGARD, New Advances in Mission Planning and Rehearsal Systems 12*, 1993.
- [24] C. A. Miller and M. D. Hannen, "The Rotorcraft Pilot's Associate: Design and evaluation of an intelligent user interface for cockpit information management," *Knowl.-Based Syst.*, vol. 12, no. 8, pp. 443–456, Dec. 1999.
- [25] P. Urlings and L. C. Jain, "Teaming human and machine: A conceptual framework," in *Hybrid Information Systems*. Heidelberg, Germany: Springer, 2002, pp. 711–721.
- [26] H. M. Bloom and N. Christopher, "A framework for distributed and virtual discrete part manufacturing," in *Proc. CALS EXPO*, 1996, pp. 1–7.
- [27] L. Vig and J. A. Adams, "A framework for multi-robot coalition formation," in *Proc. Indian Int. Conf. Artif. Intell. (IICAI)*, 2005, pp. 347–363.
- [28] J.-M. Hoc, "From human-machine interaction to human-machine cooperation," *Ergonomics*, vol. 43, no. 7, pp. 833–843, 2000.
- [29] E. Salas, C. Prince, D. P. Baker, and L. Shrestha, "Situation awareness in team performance: Implications for measurement and training," *Human Factors*, vol. 37, no. 1, pp. 123–136, 1995.
- [30] A. K. Das, R. Fierro, V. Kumar, J. P. Ostrowski, J. Spletzer, and C. J. Taylor, "A vision-based formation control framework," *IEEE Trans. Robot. Autom.*, vol. 18, no. 5, pp. 813–825, Oct. 2002.
- [31] T. Fong, C. Thorpe, and C. Baur, "Collaborative control: A robot-centric model for vehicle teleoperation," *Robot. Inst.*, Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-RI-TR-01-34, 2001.
- [32] M. Johnson, J. M. Bradshaw, P. J. Feltovich, C. Jonker, M. Sierhuis, and B. van Riemsdijk, "Toward coactivity," in *Proc. 5th ACM/IEEE Int. Conf. Human-Robot Interact.*, Mar. 2010, pp. 101–102.
- [33] M. Johnson et al., "Beyond cooperative robotics: The central role of interdependence in coactive design," *IEEE Intell. Syst.*, vol. 26, no. 3, pp. 81–88, May 2011.
- [34] E. DeKoven and A. K. Murphy, *A Framework for Supporting Teamwork Between Humans and Autonomous Systems*, DTIC document ADA463300, 2006.
- [35] J. M. Bradshaw, P. J. Feltovich, H. Jung, S. Kulkarni, W. Taysom, and A. Uszok, "Dimensions of adjustable autonomy and mixed-initiative interaction," in *Proc. Int. Workshop Comput. Auto.* Springer, 2003, pp. 17–39.
- [36] T. T. Hewett et al., *ACM SIGCHI Curricula for Human-Computer Interaction*. New York, NY, USA: ACM, 1992.
- [37] W. S. Bainbridge, *Berkshire Encyclopedia of Human-Computer Interaction*. Great Barrington, MA, USA: Berkshire Publishing Group LLC, 2004.
- [38] B. D. Argall and A. G. Billard, "A survey of tactile human-robot interactions," *Robot. Auto. Syst.*, vol. 58, no. 10, pp. 1159–1176, 2010.
- [39] J. Bray-Miners, C. Ste-Croix, and A. Morton, *Human-Robot Interaction Literature Review*, DTIC document ADA592508, 2012.
- [40] J. Cannan and H. Hu, *Human-Machine Interaction (HMI): A Survey*. Colchester, U.K.: Univ. Essex, 2011.
- [41] J. C. R. Licklider, "Man-computer symbiosis," *IRE Trans. Human Factors Electron.*, vol. HFE-1, no. 1, pp. 4–11, Mar. 1960.
- [42] R. Packer and K. Jordan, *Multimedia: From Wagner to Virtual Reality*. New York, NY, USA: Norton, 2002.
- [43] M. A. Valentine and A. C. Edmondson, "Team scaffolds: How mesolevel structures enable role-based coordination in temporary groups," *Org. Sci.*, vol. 26, no. 2, pp. 405–422, 2014.
- [44] K. J. Klein, J. C. Ziegert, A. P. Knight, and Y. Xiao, "Dynamic delegation: Shared, hierarchical, and deindividualized leadership in extreme action teams," *Administ. Sci. Quart.*, vol. 51, no. 4, pp. 590–621, 2006.
- [45] G. A. Bigley and K. H. Roberts, "The incident command system: High-reliability organizing for complex and volatile task environments," *Acad. Manage. J.*, vol. 44, no. 6, pp. 1281–1299, 2001.
- [46] K. Mackey, "Stages of team development," *IEEE Softw.*, vol. 16, no. 4, pp. 90–91, Jul. 1999.
- [47] M. H. Immordino-Yang and A. Damasio, "We feel, therefore we learn: The relevance of affective and social neuroscience to education," *Mind, Brain, Educ.*, vol. 1, no. 1, pp. 3–10, 2007.
- [48] M. Farrell, M. Schmitt, and G. Heinemann, "Informal roles and the stages of interdisciplinary team development," *J. Interprof. Care*, vol. 15, no. 3, pp. 281–295, 2001.
- [49] D. L. Miller, "The stages of group development: A retrospective study of dynamic team processes," *Can. J. Administ. Sci./Revue Canadienne des Sci. l'Admin.*, vol. 20, no. 2, pp. 121–134, 2003.
- [50] B. W. Tuckman and M. A. C. Jensen, "Stages of small-group development revisited," *Group Facilitation*, vol. 2, no. 4, pp. 419–427, 1977.
- [51] M. A. Marks, J. E. Mathieu, and S. J. Zaccaro, "A temporally based framework and taxonomy of team processes," *Acad. Manage. Rev.*, vol. 26, no. 3, pp. 356–376, 2001.
- [52] N. Mohr and A. Dichter, *Stages of Team Development*. New York, NY, USA: McGraw-Hill, 1996.
- [53] M. R. Endsley, "Design and evaluation for situation awareness enhancement," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, 1988, vol. 32, no. 2, Newbury Park, CA, USA: Sage, pp. 97–101.
- [54] N. Callow, M. J. Smith, L. Hardy, C. A. Arthur, and J. Hardy, "Measurement of transformational leadership and its relationship with team cohesion and performance level," *J. Appl. Sport Psychol.*, vol. 21, no. 4, pp. 395–412, 2009.
- [55] S. Jacklin, J. Schumann, P. Gupta, M. Richard, K. Guenther, and F. Soares, "Development of advanced verification and validation procedures and tools for the certification of learning systems in aerospace applications," in *Proc. Infotech@Aerospace*, 2005, p. 6912.

- [56] J.-P. Heuzé, N. Raimbault, and P. Fontayne, "Relationships between cohesion, collective efficacy and performance in professional basketball teams: An examination of mediating effects," *J. Sports Sci.*, vol. 24, no. 1, pp. 59–68, 2006.
- [57] D. S. Wilson, E. Ostrom, and M. E. Cox, "Generalizing the core design principles for the efficacy of groups," *J. Econ. Behav. Org.*, vol. 90, pp. S21–S32, Jun. 2013.
- [58] M. Johnson, C. Jonker, B. van Riemsdijk, P. J. Feltovich, and J. M. Bradshaw, "Joint activity testbed: Blocks world for teams (BW4T)," in *Proc. Int. Workshop Eng. Soc. Agents World*. Berlin, Germany: Springer, 2009, pp. 254–256.
- [59] D. Calisi, L. Iocchi, and D. Nardi, "A unified benchmark framework for autonomous Mobile robots and Vehicles Motion Algorithms (MoVeMA benchmarks)," in *Proc. Workshop Exp. Methodol. Benchmarking Robot. Res. (RSS)*, 2008, pp. 1–4.
- [60] R. R. Murphy and D. Schreckenghost, "Survey of metrics for human-robot interaction," in *Proc. 8th ACM/IEEE Int. Conf. Human-Robot Interact. (HRI)*, Mar. 2013, pp. 197–198.
- [61] M. L. Cummings, P. E. Pina, and B. Donmez, "Selecting metrics to evaluate human supervisory control applications," Dept. Aeronaut. Astronaut., Humans Automat. Lab., Massachusetts Inst. Technol., Boston, MA, USA, Tech. Rep. HAL2008-04, 2008.
- [62] A. Steinfeld et al., "Common metrics for human-robot interaction," in *Proc. 1st ACM SIGCHI/SIGART Conf. Human-Robot Interact.*, 2006, pp. 33–40.
- [63] M. R. Elara, C. A. A. Calderon, C. Zhou, and W. S. Wijesoma, "False alarm metrics: Evaluating safety in human robot interactions," in *Proc. IEEE Conf. Robot. Automat. Mechatronics (RAM)*, Jun. 2010, pp. 230–236.
- [64] R. E. Mohan, W. S. Wijesoma, C. A. A. Calderon, and C. Zhou, "False alarm metrics for human-robot interactions in service robots," *Adv. Robot.*, vol. 24, no. 13, pp. 1841–1859, 2010.
- [65] D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, "Teleoperation of multiple social robots," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 42, no. 3, pp. 530–544, May 2012.
- [66] T. A. O'Connell and Y.-Y. Choong, "Metrics for measuring human interaction with interactive visualizations for information analysis," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2008, pp. 1493–1496.
- [67] G. Sammer et al., "Relationship between regional hemodynamic activity and simultaneously recorded EEG-theta associated with mental arithmetic-induced workload," *Human Brain Mapping*, vol. 28, no. 8, pp. 793–803, 2007.
- [68] J. W. Crandall and M. L. Cummings, "Developing performance metrics for the supervisory control of multiple robots," in *Proc. ACM/IEEE Int. Conf. Human-Robot Interact.*, Mar. 2007, pp. 33–40.
- [69] J. Y. C. Chen and M. J. Barnes, "Human-agent teaming for multirobot control: A review of human factors issues," *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 1, pp. 13–29, Feb. 2014.
- [70] L. Rapolienė, A. Razbadauskas, J. Šalyga, and A. Martinkėnas, "Stress and fatigue management using balneotherapy in a short-time randomized controlled trial," *Evidence-Based Complementary Alternative Med.*, vol. 2016, Feb. 2016, Art. no. 9631684, doi: [10.1155/2016/9631684](https://doi.org/10.1155/2016/9631684).
- [71] T. Yazawa et al., "Measurement of stress using DFA heartbeat analysis," in *Proc. World Congr. Eng. Comput. Sci.*, vol. 2, 2013, pp. 596–603.
- [72] X. Corbillion, G. Simon, A. Devlic, and J. Chakareski, "Viewport-adaptive navigable 360-degree video delivery," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–7.
- [73] N. Forouzandehmehr, S. M. Perlaza, Z. Han, and H. V. Poor, "A satisfaction game for heating, ventilation and air conditioning control of smart buildings," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 3164–3169.
- [74] W. Saad, A. L. Glass, N. B. Mandayam, and H. V. Poor, "Toward a consumer-centric grid: A behavioral perspective," *Proc. IEEE*, vol. 104, no. 4, pp. 865–882, Apr. 2016.
- [75] W. L. Chappelle, K. McDonald, and R. E. King, "Psychological attributes critical to the performance of MQ-1 predator and MQ-9 reaper U.S. Air Force sensor operators," Air Force Res. Lab., San Antonio, TX, USA, Tech. Rep. AFRL-SA-BR-TR-2010-0007, 2010.
- [76] W. Chappelle, K. McDonald, and K. McMillan, "Important and critical psychological attributes of USAF MQ-1 predator MQ-9 reaper pilots according to subject matter experts," USAF School Aerosp. Med., Wright Patterson Air Force Base, OH, USA, Tech. Rep. AFRL-SA-WP-TR-2011-0002, 2011.
- [77] D. R. Hunter and J. E. Stewart, *Locus of Control, Risk Orientation, and Decision Making Among U.S. Army Aviators*, DTIC document ADA509824, 2009.
- [78] G. Matthews, J. S. Warm, T. H. Shaw, and V. S. Finomore, "Predicting battlefield vigilance: A multivariate approach to assessment of attentional resources," *Ergonomics*, vol. 57, no. 6, pp. 856–875, 2014.
- [79] A. Gorbenco, V. Popov, and A. Sheka, "Robot self-awareness: Exploration of internal states," *Appl. Math. Sci.*, vol. 6, no. 14, pp. 675–688, 2012.
- [80] D. Schreckenghost, T. Milam, and T. Fong, "Measuring performance in real time during remote human-robot operations with adjustable autonomy," *IEEE Intell. Syst.*, vol. 25, no. 5, pp. 36–45, Sep./Oct. 2010.
- [81] J. A. Saleh and F. Karray, "Towards generalized performance metrics for human-robot interaction," in *Proc. Int. Conf. Auto. Intell. Syst. (AIS)*, Jun. 2010, pp. 1–6.
- [82] J. W. Crandall, M. A. Goodrich, D. R. Olsen, and C. W. Nielsen, "Validating human-robot interaction schemes in multitasking environments," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 35, no. 4, pp. 438–449, Jul. 2005.
- [83] M. Chen, M. Mozaffari, W. Saad, C. Yin, M. Debbah, and C. S. Hong, "Caching in the sky: Proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1046–1061, May 2017.
- [84] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Unmanned aerial vehicle with underlaid device-to-device communications: Performance and tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 3949–3963, Jun. 2016.
- [85] E. Zeydan et al., "Big data caching for networking: Moving from cloud to edge," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 36–42, Sep. 2016.
- [86] J. Tadrus, A. Eryilmaz, and H. El Gamal, "Proactive resource allocation: Harnessing the diversity and multicast gains," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 4833–4854, Aug. 2013.
- [87] A. Khreishah, J. Chakareski, and A. Gharaibeh, "Joint caching, routing, and channel assignment for collaborative small-cell cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2275–2284, Aug. 2016.
- [88] S. J. Zaccaro, A. L. Rittman, and M. A. Marks, "Team leadership," *Leadership Quart.*, vol. 12, no. 4, pp. 451–483, 2001.
- [89] S.-J. Chen and L. Lin, "Modeling team member characteristics for the formation of a multifunctional team in concurrent engineering," *IEEE Trans. Eng. Manag.*, vol. 51, no. 2, pp. 111–124, May 2004.
- [90] F. L. Greitzer, "Toward the development of cognitive task difficulty metrics to support intelligence analysis research," in *Proc. 4th IEEE Conf. Cognit. Inform. (ICCI)*, Aug. 2005, pp. 315–320.
- [91] J. Meyer, "Effects of warning validity and proximity on responses to warnings," *Human Factors*, vol. 43, no. 4, pp. 563–572, 2001.
- [92] G. Doisy, J. Meyer, and Y. Edan, "The impact of human-robot interface design on the use of a learning robot system," *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 6, pp. 788–795, Dec. 2014.
- [93] J. A. Shah, J. H. Saleh, and J. A. Hoffman, "Review and synthesis of considerations in architecting heterogeneous teams of humans and robots for optimal space exploration," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 37, no. 5, pp. 779–793, Sep. 2007.
- [94] P. Walker, S. Nunnally, M. Lewis, A. Kolling, N. Chakraborty, and K. Sycara, "Neglect benevolence in human control of swarms in the presence of latency," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2012, pp. 3009–3014.
- [95] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén, "Measurement," in *Experimentation in Software Engineering*. Berlin, Germany: Springer, 2012, pp. 37–43, doi: [10.1007/978-3-642-29044-2](https://doi.org/10.1007/978-3-642-29044-2).
- [96] P. Pina, M. L. Cummings, J. W. Crandall, and M. D. Penna, "Identifying generalizable metric classes to evaluate human-robot teams," in *Proc. 3rd Ann. Conf. Human-Robot Interact.*, 2008, pp. 13–20.
- [97] M. A. Goodrich, S. Kerman, and S.-Y. Jun, "On leadership and influence in human-swarm interaction," in *Proc. AAAI Fall Symp., Human Control Bioinspired Swarms*, 2012, pp. 1–6.
- [98] M. A. Goodrich, B. Pendleton, P. B. Sujit, J. Pinto, and J. W. Crandall, "Toward human interaction with bio-inspired teams," in *Proc. 10th Int. Conf. Auto. Agents Multiagent Syst.*, vol. 3, 2011, pp. 1265–1266.
- [99] J. D. Lee, "Emerging challenges in cognitive ergonomics: Managing swarms of self-organizing agent-based automation," *Theor. Issues Ergonom. Sci.*, vol. 2, no. 3, pp. 238–250, 2001.

- [100] E. Ferrante, A. E. Turgut, C. Huepe, A. Stranieri, C. Pincioli, and M. Dorigo, "Self-organized flocking with a mobile robot swarm: A novel motion control method," *Adapt. Behav.*, vol. 20, no. 6, pp. 460–477, 2012.
- [101] W. M. Spears, D. F. Spears, J. C. Hamann, and R. Heil, "Distributed, physics-based control of swarms of vehicles," *Auto. Robots*, vol. 17, no. 2, pp. 137–162, Sep. 2004.
- [102] M. D. Mumford, W. A. Baughman, K. V. Threlfall, C. E. Uhlman, and D. P. Costanza, "Personality, adaptability, and performance: Performance on well-defined problem solving tasks," *Human Perform.*, vol. 6, no. 3, pp. 241–285, 1993.
- [103] S. A. Rathus, "A 30-item schedule for assessing assertive behavior," *Behav. Therapy*, vol. 4, no. 3, pp. 398–406, May 1973.
- [104] M. J. Pearsall, A. P. J. Ellis, and M. J. Pearsall, "The effects of critical team member assertiveness on team performance and satisfaction," *J. Manage.*, vol. 32, no. 4, pp. 575–594, Aug. 2006.
- [105] A. J. Martin and H. W. Marsh, "Academic resilience and the four Cs: Confidence, control, composure, and commitment," in *Proc. NZARE Austral. Assoc. Res. Educ. (AARE)*, Auckland, New Zealand, 2003, p. 12. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.86.5636&rep=rep1&type=pdf>
- [106] G. Windle, K. M. Bennett, and J. Noyes, "A methodological review of resilience measurement scales," *Health Quality Life Outcomes*, vol. 9, no. 1, p. 8, 2011.
- [107] J. A. Green, D. B. O'Connor, N. Gartland, and B. W. Roberts, "The Chernysenko conscientiousness scales: A new facet measure of conscientiousness," *Assessment*, vol. 23, no. 3, pp. 374–385, 2016.
- [108] V. Dulewicz and M. Higgs, "Can emotional intelligence be measured and developed?" *Leadership Org. Develop. J.*, vol. 20, no. 5, pp. 242–253, 1999.
- [109] J. Wilt and W. Revelle, "Extraversion," Tech. Rep., 2008. Following is more details about citation: J. Wilt and W. Revelle, "Extraversion," in *Handbook of Individual Differences in Social Behavior*, M. R. Leary and R. H. Hoyle, Eds. New York, NY, USA: Guilford Press, 2009, pp. 27–45.
- [110] T. Heffernan and J. Ling, "The impact of Eysenck's extraversion-introversion personality dimension on prospective memory," *Scand. J. Psychol.*, vol. 42, no. 4, pp. 321–325, 2001.
- [111] I. B. Mauss and M. D. Robinson, "Measures of emotion: A review," *Cognit. Emotion*, vol. 23, no. 2, pp. 209–237, 2009.
- [112] R. L. Tomko et al., "Measuring impulsivity in daily life: The momentary impulsivity scale," *Psychol. Assessment*, vol. 26, no. 2, p. 339, 2014.
- [113] M. R. Endsley, "Measurement of situation awareness in dynamic systems," *Human Factors*, vol. 37, no. 1, pp. 65–84, 1995.
- [114] R. Christensen and G. Knezek, "Comparative measures of grit, tenacity and perseverance," *Int. J. Learn., Teach. Educ. Res.*, vol. 8, no. 1, pp. 16–30, 2014.
- [115] K. Parsons, A. McCormac, M. Butavicius, M. Pattinson, and C. Jerram, "Determining employee awareness using the Human Aspects of Information Security Questionnaire (HAIS-Q)," *Comput. Secur.*, vol. 42, pp. 165–176, May 2014.
- [116] S. G. Hart, "Nasa-task load index (NASA-TLX): 20 years later," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, 2006, vol. 50, no. 9. Los Angeles, CA, USA: Sage, pp. 904–908.
- [117] P. T. Costa and R. R. McCrae, "Neuroticism, somatic complaints, and disease: Is the bark worse than the bite?" *J. Pers.*, vol. 55, no. 2, pp. 299–316, 1987.
- [118] X. Fan et al., "An exploratory study about inaccuracy and invalidity in adolescent self-report surveys," *Field Methods*, vol. 18, no. 3, pp. 223–244, 2006.
- [119] R. Wilcox, *Introduction to Robust Estimation and Hypothesis Testing*. San Francisco, CA, USA: Academic, 2011.
- [120] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.
- [121] J. E. Ware, Jr., "Scales for measuring general health perceptions," *Health Services Res.*, vol. 11, no. 4, p. 396, 1976.
- [122] E. Åhsberg, "Dimensions of fatigue in different working populations," *Scand. J. Psychol.*, vol. 41, no. 3, pp. 231–241, 2000.
- [123] S. J. Lupien and F. Seguin, "How to measure stress in humans," in *Centre for Studies in Human Stress*. Montréal, QC, Canada: Fernand-Seguin Research Center Louis-H. Lafontaine Hospital, 2013.
- [124] M. J. Barnes and M. F. Matz, "Crew simulations for unmanned aerial vehicle (UAV) applications: Sustained effects, shift factors, interface issues, and crew size," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, Oct. 1998, vol. 42, no. 1, pp. 143–147.
- [125] S. Folkard and P. Tucker, "Shift work, safety and productivity," *Occupat. Med.*, vol. 53, no. 2, pp. 95–101, 2003.
- [126] J. Meyer, Y. Bitan, D. Shinar, and E. Zmora, "Scheduling of actions and reliance on warnings in a simulated control task," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, 1999, vol. 43, no. 3. Los Angeles, CA, USA: Sage, pp. 251–255.
- [127] E. W. Russell, "A multiple scoring method for the assessment of complex memory functions," *J. Consulting Clin. Psychol.*, vol. 43, no. 6, pp. 800–809, 1975.
- [128] S. E. Devena and M. W. Watkins, "Diagnostic utility of WISC-IV general abilities index and cognitive proficiency index difference scores among children with ADHD," *J. Appl. School Psychol.*, vol. 28, no. 2, pp. 133–154, 2012.
- [129] National Research Council, *Complex Operational Decision Making in Networked Systems of Humans and Machines: A Multidisciplinary Approach*. Washington, DC, USA: National Academies Press, 2014.
- [130] K. G. Shin and P. Ramanathan, "Real-time computing: A new discipline of computer science and engineering," *Proc. IEEE*, vol. 82, no. 1, pp. 6–24, Jan. 1994.
- [131] V. Groom and C. Nass, "Can robots be teammates?: Benchmarks in human-robot teams," *Interact. Stud.*, vol. 8, no. 3, pp. 483–500, 2007.
- [132] D. R. Olsen, Jr., and M. A. Goodrich, "Metrics for evaluating human-robot interactions," in *Proc. PERMIS*, 2003, p. 4.
- [133] M. H. Smaili, J. Breeman, T. J. J. Lombaerts, and O. Stroosma, "A simulation benchmark for aircraft survivability assessment," in *Proc. 26th Int. Congr. Aeronaut. Sci.*, Anchorage, AK, USA, 2008, vol. 9, no. 2, pp. 1–12.
- [134] J. Annett, "Subjective rating scales: Science or art?" *Ergonomics*, vol. 45, no. 14, pp. 966–987, 2002.
- [135] L. M. Naismith, J. J. H. Cheung, C. Ringsted, and R. B. Cavalcanti, "Limitations of subjective cognitive load measures in simulation-based procedural training," *Med. Educ.*, vol. 49, no. 8, pp. 805–814, 2015.
- [136] J. C. Bol and S. D. Smith, "Spillover effects in subjective performance evaluation: Bias and the asymmetric influence of controllability," *Accounting Rev.*, vol. 86, no. 4, pp. 1213–1230, 2011.
- [137] J. W. Crandall and M. L. Cummings, "Identifying predictive metrics for supervisory control of multiple robots," *IEEE Trans. Robot.*, vol. 23, no. 5, pp. 942–951, Oct. 2007.
- [138] J. A. Weekley and R. E. Ployhart, Eds., *Situational Judgment Tests: Theory, Measurement, and Application*. London, U.K.: Psychology Press, 2013.
- [139] D. Dowding and C. Thompson, "Measuring the quality of judgement and decision-making in nursing," *J. Adv. Nursing*, vol. 44, no. 1, pp. 49–57, 2003.
- [140] M. A. McDaniel, F. P. Morgeson, E. B. Finnegan, M. A. Campion, and E. P. Braverman, *Use of Situational Judgment Tests to Predict Job Performance: A Clarification of the Literature*. Washington, DC, USA: American Psychological Association, 2001.
- [141] F. Patterson, V. Ashworth, L. Zibarras, P. Coan, M. Kerrin, and P. O'Neill, "Evaluations of situational judgement tests to assess non-academic attributes in selection," *Med. Educ.*, vol. 46, no. 9, pp. 850–868, 2012.
- [142] J. Clevenger, G. M. Pereira, D. Wiechmann, N. Schmitt, and V. S. Harvey, "Incremental validity of situational judgment tests," *J. Appl. Psychol.*, vol. 86, no. 3, pp. 410–417, 2001.
- [143] K. M. Kowalski-Trakofler, C. Vaught, and T. Scharf, "Judgment and decision making under stress: An overview for emergency managers," *Int. J. Emergency Manage.*, vol. 1, no. 3, pp. 278–289, 2003.
- [144] H. F. Chau, C. Y. Wong, F. K. Chow, and C.-H. F. Fung, "Social judgment theory based model on opinion formation, polarization and evolution," *Phys. A, Statist. Mech. Appl.*, vol. 415, pp. 133–140, Dec. 2014.
- [145] L. A. Rabin et al., "Judgment in older adults: Development and psychometric evaluation of the Test of Practical Judgment (TOP-J)," *J. Clin. Exp. Neuropsychol.*, vol. 29, no. 7, pp. 752–767, 2007.
- [146] A. S. Clare, M. L. Cummings, and N. P. Repenning, "Influencing trust for human-automation collaborative scheduling of multiple unmanned vehicles," *Human Factors*, vol. 57, no. 7, pp. 1208–1218, 2015.
- [147] S. Bruni, J. J. Marquez, A. Brzezinski, and M. L. Cummings, "Visualizing operators' cognitive strategies in multivariate optimization," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, 2006, vol. 50, no. 11. Los Angeles, CA, USA: Sage, pp. 1180–1184.
- [148] S. Heuer and B. Hallowell, "A novel eye-tracking method to assess attention allocation in individuals with and without aphasia using a dual-task paradigm," *J. Commun. Disorders*, vol. 55, pp. 15–30, May/Jun. 2015.

- [149] T. Roderer and C. M. Roebers, "Can you see me thinking (about my answers)? Using eye-tracking to illuminate developmental differences in monitoring and control skills and their relation to performance," *Metacognit. Learn.*, vol. 9, no. 1, pp. 1–23, 2014.
- [150] N. Moacdieh and N. Sarter, "Clutter in electronic medical records: Examining its performance and attentional costs using eye tracking," *Human Factors*, vol. 57, no. 4, pp. 591–606, 2015.
- [151] P. Damacharla, A. Y. Javaid, and V. K. Devabhaktuni, "Human error prediction using eye tracking to improvise team cohesion in human-machine teams," in *Proc. 9th Int. Conf. Appl. Human Factors Ergonom. (AHFE)*, Orlando, FL, USA: Springer, 2018, pp. 47–57.
- [152] C. R. Kovesdi, B. C. Rice, G. R. Bower, Z. A. Spielman, R. A. Hill, and K. L. LeBlanc, "Measuring human performance in simulated nuclear power plant control rooms using eye tracking," Idaho Nat. Lab, Idaho Falls, ID, USA, Tech. Rep. INL/EXT-15-37311, 2015.
- [153] F. Paas, J. E. Tuovinen, H. Tabbers, and P. W. M. Van Gerven, "Cognitive load measurement as a means to advance cognitive load theory," *Educ. Psychol.*, vol. 38, no. 1, pp. 63–71, 2003.
- [154] F. A. Haji, D. Rojas, R. Childs, S. Ribaupierre, and A. Dubrowski, "Measuring cognitive load: Performance, mental effort and simulation task complexity," *Med. Educ.*, vol. 49, no. 8, pp. 815–827, 2015.
- [155] T. A. Nguyen and Y. Zeng, "A physiological study of relationship between designer's mental effort and mental stress during conceptual design," *Comput.-Aided Des.*, vol. 54, pp. 3–18, Sep. 2014.
- [156] G. Matthews, L. E. Reinerman-Jones, D. J. Barber, and J. Abich, "The psychometrics of mental workload: Multiple measures are sensitive but divergent," *Human Factors*, vol. 57, no. 1, pp. 125–143, 2014.
- [157] T. Fritz, A. Begel, S. C. Müller, S. Yigit-Elliott, and M. Züger, "Using psycho-physiological measures to assess task difficulty in software development," in *Proc. 36th Int. Conf. Softw. Eng.*, 2014, pp. 402–413.
- [158] T. Glenn and S. Monteith, "New measures of mental state and behavior based on data collected from sensors, smartphones, and the Internet," *Current Psychiatry Rep.*, vol. 16, no. 12, p. 523, 2014.
- [159] M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, "State of science: Mental workload in ergonomics," *Ergonomics*, vol. 58, no. 1, pp. 1–17, 2015.
- [160] A. Andre and A. Degani, "Do you know what mode you're in? An analysis of mode error in everyday things," in *Human-Automation Interaction: Research and Practice*. Mahwah, NJ, USA: Lawrence Erlbaum, 1997, pp. 19–28.
- [161] N. B. Sarter and D. D. Woods, "Mode error in supervisory control of automated systems," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, 1992, vol. 36, no. 1. Los Angeles, CA, USA: Sage, pp. 26–29.
- [162] M. R. Elara, C. A. A. Calderón, C. Zhou, and W. S. Wijesoma, "Experimenting extended neglect tolerance model for human robot interactions in service missions," in *Proc. 11th Int. Conf. Control Automat. Robot. Vis. (ICARCV)*, Dec. 2010, pp. 2024–2029.
- [163] J. A. Saleh, F. Karay, and M. Morckos, "Modelling of robot attention demand in human-robot interaction using finite fuzzy state automata," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jun. 2012, pp. 1–8.
- [164] J. W. Crandall, C. W. Nielsen, and M. A. Goodrich, "Towards predicting robot team performance," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, vol. 1, Oct. 2003, pp. 906–911.
- [165] T. W. McLain, P. R. Chandler, S. Rasmussen, and M. Pachter, "Cooperative control of UAV rendezvous," in *Proc. Amer. Control Conf.*, vol. 3, Jun. 2001, pp. 2309–2314.
- [166] L. Berardinelli, V. Cortellessa, and A. Di Marco, "Performance modeling and analysis of context-aware mobile software systems," in *Proc. Int. Conf. Fundam. Approaches Softw. Eng.*, 2010, pp. 353–367.
- [167] D. Harel, "Statecharts: A visual formalism for complex systems," *Sci. Comput. Program.*, vol. 8, no. 3, pp. 231–274, 1987.
- [168] P. Damacharla, R. R. Junuthula, A. Y. Javaid, and V. K. Devabhaktuni, "Autonomous ground vehicle error prediction modeling to facilitate human-machine cooperation," in *Proc. 9th Int. Conf. Appl. Human Factors Ergonom. (AHFE)*, Orlando, FL, USA: Springer, 2018, pp. 36–45.
- [169] T. Schlogl, C. Beleznaï, M. Winter, and H. Bischof, "Performance evaluation metrics for motion detection and tracking," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, vol. 4, Aug. 2004, pp. 519–522.
- [170] F. Serban, J. Vanschoren, J.-U. Kietz, and A. Bernstein, "A survey of intelligent assistants for data analysis," *ACM Comput. Surv.*, vol. 45, no. 3, p. 31, 2013.
- [171] J. S. Ash, M. Berg, and E. Coiera, "Some unintended consequences of information technology in health care: The nature of patient care information system-related errors," *J. Amer. Med. Inform. Assoc.*, vol. 11, no. 2, pp. 104–112, 2004.
- [172] T. Peynot, J. Underwood, and S. Scheduling, "Towards reliable perception for unmanned ground vehicles in challenging conditions," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, St. Louis, MO, USA, Oct. 2009, pp. 1170–1176.
- [173] S. C. Peters and K. Iagnemma, "An analysis of rollover stability measurement for high-speed mobile robots," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, Orlando, FL, USA, May 2006, pp. 3711–3716.
- [174] W. Burgard et al., "Experiences with an interactive museum tour-guide robot," *Artif. Intell.*, vol. 114, nos. 1–2, pp. 3–55, 1999.
- [175] D. Schreckenghost, T. Fong, H. Utz, and T. Milam, "Measuring robot performance in real-time for NASA robotic reconnaissance operations," in *Proc. 9th Workshop Perform. Metrics Intell. Syst.*, 2009, pp. 194–202.
- [176] A. V. Carron and L. R. Brawley, "Cohesion: Conceptual and measurement issues," *Small Group Res.*, vol. 31, no. 1, pp. 89–106, 2000.
- [177] A. V. Carron and J. R. Ball, "An analysis of the cause-effect characteristics of cohesiveness and participation motivation in intercollegiate hockey," *Int. Rev. Sport Sociol.*, vol. 12, no. 2, pp. 49–60, 1977.
- [178] E. Salas, R. Grossman, A. M. Hughes, and C. W. Coultas, "Measuring team cohesion: Observations from the Science," *Human Factors*, vol. 57, no. 3, pp. 365–374, 2015.
- [179] J. Burke and R. Murphy, "RSVP: An investigation of remote shared visual presence as common ground for human-robot teams," in *Proc. 2nd ACM/IEEE Int. Conf. Human-Robot Interact. (HRI)*, Mar. 2007, pp. 161–168.
- [180] S.-J. Yi et al., "Team THOR's entry in the DARPA robotics challenge trials 2013," *J. Field Robot.*, vol. 32, no. 3, pp. 315–335, 2015.
- [181] H.-M. Huang, E. Messina, and J. Albus, "Toward a generic model for autonomy levels for unmanned systems (ALFUS)," Intell. Syst. Division, Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Tech. Rep. ADA515323, 2003.
- [182] Y. Nakamura, H. Hanafusa, and T. Yoshikawa, "Task-priority based redundancy control of robot manipulators," *Int. J. Robot. Res.*, vol. 6, no. 2, pp. 3–15, 1987.
- [183] V. Klamroth-Marganska et al., "Three-dimensional, task-specific robot therapy of the arm after stroke: A multicentre, parallel-group randomised trial," *Lancet Neurol.*, vol. 13, no. 2, pp. 159–166, 2014.
- [184] P. Damacharla et al., "Effects of voice-based synthetic assistant on performance of emergency care provider in training," *Int. J. Artif. Intell. Educ.*, pp. 1–22, Mar. 2018. [Online]. Available: <https://link.springer.com/article/10.1007/s40593-018-0166-3>
- [185] S. Prabhala, J. J. Gallimore, and J. R. Lucas, "Evaluating human interaction with automation in a complex UCAV control station simulation using multiple performance metrics," in *Human-in-the-Loop Simulations*. London, U.K.: Springer, 2011, pp. 239–258.
- [186] J. Gerhardt-Powals, "Cognitive engineering principles for enhancing human-computer performance," *Int. J. Human-Comput. Interact.*, vol. 8, no. 2, pp. 189–211, 1996.
- [187] H. Cai and Y. Lin, "Modeling of operators' emotion and task performance in a virtual driving environment," *Int. J. Human-Comput. Stud.*, vol. 69, no. 9, pp. 571–586, Aug. 2011.
- [188] A. Steinfeld, "Interface lessons for fully and semi-autonomous mobile robots," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, vol. 3, Apr./May 2004, pp. 2752–2757.



PRAVEEN DAMACHARLA (S'11–GS'13) received the B.Tech. degree in electrical and electronics engineering from the Koneru Lakshmaiah College of Engineering, Acharya Nagarjuna University, Guntur, India, in 2012. He is currently pursuing the Ph.D. degree with The University of Toledo, Toledo, OH, USA. He is currently a Research Assistant with the Department of Electrical Engineering and Computer Science, The University of Toledo. His research interests include human-machine teaming, human factors, machine learning, autonomous synthetic assistants, and applied robotics. He was a recipient of the Outstanding Teaching Assistant 2015 Award presented by the College of Engineering, The University of Toledo. He was also a recipient of the 2016 Advanced Leadership Academy Scholarship from the College of Business and Innovation, The University of Toledo.



AHMAD Y. JAVAID (GS'12–M'15) received the B.Tech. degree (Hons.) in computer engineering from Aligarh Muslim University, India, in 2008, the Ph.D. degree from The University of Toledo (UT) in 2015 along with the prestigious University Fellowship Award. Previously, he worked for two years as a Scientist Fellow with the Ministry of Science & Technology, Government of India. He joined the EECS Department as an Assistant Professor in Fall 2015 and

is the Founding Director of the Paul A. Hotmer Cybersecurity and Teaming Research Lab, UT. He is also conducting extensive research on human-machine teams and applications of AI and machine learning to attack detection and mitigation. During his time at UT, he has participated in several collaborative research proposals that have been funded by agencies, including the National Science Foundation, Air Force Research Lab, and the State of Ohio. He has published over 50 peer-reviewed journal, conference, and poster papers along with several book chapters. His research expertise is in the area of cybersecurity of drone networks, smartphones, wireless sensor networks, and other systems. He has also served as a reviewer for several high impact IEEE journals and as a member of the technical program committee for several conferences.



JENNIE J. GALLIMORE received the master's degree in psychology and the Ph.D. degree in industrial engineering and operations research from the Virginia Polytechnic Institute and State University in 1985 and 1989, respectively. She was a Professor of industrial and human factors engineering with the Department of Biomedical, Industrial and Human Factors Engineering, Wright State University. She used to hold a joint appointment as a Professor with the Department of

Surgery, Boonshoft School of Medicine. She applies human factors and cognitive engineering principles to the design of complex systems. She is currently the Dean for the College of Technology, Architecture and Applied Engineering, Bowling Green State University. She has published over 60 technical articles and has taught 13 different courses related to human factors engineering. She has worked in the research domains of visualization of information, aviation, unmanned aerial systems, health care systems, petrochemical, and virtual environments to name a few. She was a recipient of the 2001 AIAA Simulation and Modeling Best Paper Award and the WSU College of Engineering and Computer Science Excellence in Research Award 2001/2002. Three of her Ph.D. students received the Stanley N. Roscoe Award for best dissertation in Aviation Human Factors (2000, 2003, and 2007).



VIJAY K. DEVABHAKTUNI (S'97–M'03–SM'09) received the B.Eng. degree in EEE and the M.Sc. degree in physics from BITS, Pilani, in 1996, and the Ph.D. in electronics from Carleton University, Canada, in 2003. From 2003 to 2004, he held the Natural Sciences and Engineering Research Council of the Canada Post-Doctoral Fellowship and spent the tenure researching with the University of Calgary. In 2005, he taught with Penn State Erie Behrend. From 2005 to 2008, he held

the internationally prestigious Canada Research Chair in computer-aided high-frequency modeling and design with Concordia University, Montreal, Canada. In 2008, he joined the EECS Department, The University of Toledo, as an Associate Professor. Since 2012, he has been the Director of the College of Engineering for Interdisciplinary Research Initiatives, and has been recently promoted to a Full Professor. He has co-authored around 190 peer-reviewed papers and is supervising 13 M.S./Ph.D. students. His interests are applied electromagnetics, biomedical applications of wireless networks, computer aided design, device modeling, image processing, infrastructure monitoring, neural networks, optimization methods, power theft modeling and education, RF/microwave optimization, and virtual reality. In these areas, he secured external funding close to 5 million (sponsoring agencies, include AFOSR, CFI, ODOT, NASA, NSERC, NSF, and industry). He is a Registered Member of the Association of Professional Engineers, Geologists, and Geophysicists of Alberta. He was a recipient of the Carleton University Senate Medal for outstanding scholastic accomplishments at the Ph.D. level. He was also a recipient of several teaching excellence awards in Canada and USA. He serves as the Associate Editor of the *International Journal of RF and Microwave Computer-Aided Engineering*.

...