Towards Meaningfully Integrating Human-Autonomy Teaming in Applied Settings

Beau G. Schelble
bschelb@g.clemson.edu
Clemson University
Clemson, South Carolina

Christopher Flathmann
cflathm@g.clemson.edu
Clemson University
Clemson, South Carolina

Nathan McNeese
mcneese@g.clemson.edu
Clemson University
Clemson, South Carolina

ABSTRACT

Technological advancement goes hand in hand with economic advancement, meaning applied industries like manufacturing, medicine, and retail are set to leverage new practices like human-autonomy teams. These human-autonomy teams call for deep integration between artificial intelligence and the human workers that make up a majority of the workforce. This paper identifies the core principles of the human-autonomy teaming literature relevant to the integration of human-autonomy teams in applied contexts and research due to this large scale implementation of human-autonomy teams. A framework is built and defined from these fundamental concepts, with specific examples of its use in applied contexts and the interactions between various components of the framework. This framework can be utilized by practitioners of human-autonomy teams, allowing them to make informed decisions regarding the integration and training of human-autonomy teams.

CCS CONCEPTS

• **Human-centered computing** → **HCI theory, concepts and models**; *Social engineering (social sciences)*; • **Computing methodologies** → **Artificial intelligence**; • **Applied computing** → *Industry and manufacturing*; • **General and reference** → General conference proceedings;

KEYWORDS

theoretical framework, artificial intelligence, human-autonomy teaming, human-autonomy interaction, teaming, applied artificial intelligence

ACM Reference Format:
Beau G. Schelble, Christopher Flathmann, and Nathan McNeese. 2020. Towards Meaningfully Integrating Human-Autonomy Teaming in Applied Settings. In *Proceedings of the 8th International Conference on Human-Agent Interaction (HAI '20), November 10–13, 2020, Virtual Event, NSW, Australia.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3406499.3415077

1 INTRODUCTION

Rapid advancements in the Internet of Things (IoT), big data (data visualization & sensemaking), and artificial intelligence (AI) are encouraging the advancement of several applied industries like manufacturing, health care, and consumer services [43]. This shift is, in part, enabled and inspired by advances in AI technology and its democratization [1, 36]. As a consequence of this advancement, AI is rapidly progressing towards playing a prominent role in applied settings, utilizing other data-driven technologies to drive a safe, productive, and situationally aware system no matter the industry. Consequentially, there have been many calls to produce roadmaps, models, frameworks, and implementations of AI for industry [18], which has been met with a variety of different technically focused models [22, 33]. However, a specific gap remains in developing a theoretical model for integrating human-autonomy teams (HAT) and their AI-powered teammates into applied industry settings. This model is necessary as humans working alongside autonomous agents face many potential pitfalls if those HATs are not correctly integrated into the environment. The proposed model aims to help bridge the gap between current human teams and HATs, which share several vital differences.

Typical human teams are defined by two or more human members working together towards a common goal interdependently [35], while human-autonomy teams (HATs) have significant differences in team composition given the inclusion of an autonomous agent. That autonomous agent should not be confused with a purely automated agent, which does not qualify as an autonomous agent. Whenever using the term autonomous agent, the paper is referring to the following definition of autonomy. The autonomous agents that makeup HATs are significantly different from those agents that make up human-automation teams. Autonomous agents act intelligently deciding their own courses of action through self-government and pro-activity [8, 26]. Purely automated are defined by their inability to independently partake in activities that benefit the team without being pre-programmed to do so [30]. In addition to autonomy requirements, a team does not become a HAT unless the autonomous agent is seen as a full member of the team, performing a unique and distinct role .

These HATs are going to be utilized in a variety of settings for several reasons: (1) teams are one of the most common occurrences in many people's lives and are utilized to complete tasks in a variety of contexts [34], (2) advances made in AIs abilities [1], and (3) the efficacy of these human-autonomy teams has been demonstrated [40]. With such significant driving factors pushing industries to utilize HATs, it becomes vital for practitioners to make informed decisions when integrating these HATs in their facilities. This paper presents a framework for implementing HATs in applied settings to inform practitioners and users of HATs, complete with applied examples of the frameworks use, and a discussion of its scalability.

HAI '20, November 10–13, 2020, Virtual Event, NSW, Australia

ACM ISBN 978-1-4503-8054-6/20/11…$15.00
https://doi.org/10.1145/3406499.3415077

The current framework addresses a gap in general HAT implementation caused by AI technology rapidly reaching an appropriate level of maturity coupled with the democratization of AI development resources [1, 36]. These two developments allow HATs to be deployed in applied settings more than ever before, giving researchers the ease of access to high-level technology necessary to efficiently conduct human-AI interaction/teaming experiments that were simply unavailable before.

One of the leading examples of HATs utilization in applied settings involves Industry 4.0 (I4.0) philosophy, which can potentially implement hundreds of thousands of HATs in numerous different contexts around the world. Such a large-scale integration of AI and humans coming together as teams necessitates careful consideration of how to go about designing and implementing these special teams. This paper delivers a framework for practitioners and researchers that wish to integrate autonomous agents alongside humans to create HATs in the workplace or within the lab. This framework can help ensure efficient, safe, and productive integration of HATs throughout the spectrum of applied settings, allowing practitioners to consider each facet that impacts human-autonomy teaming, like transparency, autonomy level, reliability, and individual differences, and training.

As I4.0 acts as a leading example of applied HATs, the philosophy is a driving force of the technologies that will allow HATs to exist in a variety of other settings. Such technologies include: (1) distributed interconnected devices that create the IoT, (2) the utilization of Big Data, and (3) the utilization of AI to create more flexible and adaptable systems [9, 17, 21, 24]. These technologies are leveraged throughout the I4.0 literature to introduce cloud manufacturing, cyber-physical systems, and smart factories. Cloud manufacturing and cyber-physical systems are niche technologies in advanced manufacturing but are important example use cases for applied HATs; however, the two technologies are driven by the IoT, which plays an important role in applied HATs. The IoT involves connecting each relevant component of the operation to the internet to monitor it and collect data continuously [24]. That data is collected and used to visualize and create meaning, otherwise known as big data, which is utilized to drive things like predictive maintenance, autonomous agent training, and increased financial returns [24]. The technologies utilized for I4.0, like Big Data, the IoT, and AI, are the same technologies that will allow for the use of HATs in various other applied settings, making I4.0 a valid example to apply the proposed framework in the context of this paper.

1.1 Framework Development

Conceptual works have been noted as critical contributions to the vitality of other fields [25], with human-agent interaction (HAI) being no exception to this assertion, as shown in past papers [19, 20]. That being said, any useful theoretical framework has a basis in established literature and methodology; thus, the current paper utilizes the well-regarded methods of David Whetten [41], and Deborah MacInnis [25]. Whetten described conceptual contributions as identifying critical factors relevant to the author's goal, defining their relationship with one another, and highlighting the underlying forces that drive the selection of factors and implied relationships [41]. MacInnis further described the different types of conceptual contributions, whereas the current paper is defined as an integration. Integration contributions highlight previously published knowledge and attempt to draw connections between the different phenomena to create a novel higher-order conceptualization [25]. The current paper follows these methodologies and past theoretical HAI papers [19, 20], by identifying critical concepts relevant to applied HATs, identifying relationships, and creating a novel framework.

2 HUMAN-AUTONOMY TEAMS PLACE IN APPLIED SETTINGS

Using I4.0 as an example industry, it is clear that autonomous agents powered by AI are pivotal to the implementation of various aspects of I4.0, and that AI will be brushing shoulders with a variety of human teammates. Theoretical examples of HATs in applied I4.0 contexts involve humans and autonomous agents working together to monitor, control, and make decisions for the manufacturing facility. For example, an autonomous agent tasked with ensuring the facility does not suffer any significant downtime through predictive maintenance. This autonomous agent would work with other agents and other humans to collectively monitor and preemptively make repairs to the facility before they become significant problems [11]. The autonomous agent works by making decisions based upon data coming from the interconnected devices (IoT), while the human makes decisions based on intuition, experience, and physical awareness. Each tackles the problem in different but complementary ways, as the autonomous agent and human come together to produce results more significant than each would alone. Interactions like these are necessary to achieve the global goal of the I4.0 environment, as fruitful cooperation is required between all team members, which includes humans and autonomous agents alike [32].

In more general research studies that analyze HATs in applied settings, various contexts have been used. Settings involving air traffic control, unmanned aerial vehicle (UAV) operation, and resource allocation for emergency response [26, 37, 39]. Each teaming context that the autonomous agent was utilized differs from the last, highlighting the generalizability that HATs possess. Additionally, it has been shown that properly implemented HATs result in better performance than autonomy working alone [43]. These improvements are seen as humans, and autonomous agents are coming together to increase essential business outcomes in many industries like auto manufacturing, casino management, and disease prediction [43]. With humans and autonomous agents teaming up successfully in such varied contexts, it is clear that the addition of HATs in applied settings truly complements the human workers and business itself in a meaningful manner.

3 FRAMEWORK FOR INTEGRATING HUMAN-AUTONOMOUS AGENT TEAMS INTO APPLIED CONTEXTS

The framework is shown in Figure 1 and identifies the crucial and relevant factors that relate to HAT integration in applied settings. This framework can also serve as a guide to inform proper methodologies in human-AI interaction/teaming experiments. Each factor is detailed and related to specific applied HAT use cases and other

factors within the framework. At a high level, the framework interacts with other components in two distinct domains and one outcome factor (green). The first domain of factors are human-specific factors (dark grey), and the second is autonomous agent-specific factors (light grey) with the bi-directional transparency facet (light blue) acting as a bridge between the two teammate types. Additionally, Figure 2 identifies the order in which each factor is implemented and or assessed by applied HAT practitioners, while Figure 1 is meant to convey information and relationships. With these facets identified and related to one another, the framework can come together as a generalizable tool to inform HAT integration in various applied settings.

3.1 Effects of Individual Differences

Individual differences refer to the variability of human characteristics across individuals, such as culture, working memory, and past experiences. A range of effects for individual differences have been noted in the literature involving things like personality, culture, and prior experiences. For example, if a virtual autonomous agent has similar personality traits to the human members of the HAT, the human teammates will benefit from an increased ability to develop shared mental models [14]. Human team members' general cultural viewpoints also appear to affect trust in the autonomous agent, with horizontal collectivism and individualism being viewpoints that generally lend themselves to higher trust in autonomous systems [15]. Finally, prior experiences affect individuals who have had positive prior experiences with autonomous systems saw higher levels of trust [12, 13], and those with prior video game experience benefit from enhanced performance in tasks that involve spatial ability [3].

The consideration of individual differences comes into play when determining how and where to best implement the autonomous agents within the facility and training individuals for HATs. For example, if employees have had a positive experience with automated systems in the facility in the past, the prior experience will likely result in increased trust and a more efficient transition. Additionally, the effect seen from prior experience playing video games can be implemented in training for HATs in I4.0 settings. Many HATs will be working in spatially demanding tasks, which means increased performance can be produced if training is implemented that simulates that spatial component in a virtual environment. Additionally, the autonomous agents used in training will work with many individuals human teammates, exposing it to any number of individual differences. This autonomous agent can be specially trained to identify individual differences in team members and adapt their policy to accommodate the effects of those differences as outlined previously.

In research settings, accounting for individual differences should focus on controlling for the critical individual differences highlighted in this review. Individual differences regarding past experiences that affect spatial ability (past video game experience) and positive or negative experiences with past autonomous agents can easily be collected from a pre-task survey and controlled for in an ANCOVA if necessary. Other individual differences related to culture like nationality may be more challenging to assess. This difficulty is because culturalistic viewpoints cannot readily be assumed based on nationality, and the measures tend to take more time, requiring more extensive planning to determine if the effects of the individual difference will impact the results of the study.

3.2 Training

An essential component of integrating autonomous agents into I4.0 facilities will be training employees to work with autonomous agents and vice versa. As would be expected, training with autonomous agents before performing a team task has positive results [7]. For example, a specific type of training known as cross-training displays a great deal of success in HATs [28]. This type of training has the human and autonomous agent switch roles in order to learn more about the other's job (if they are complementary). A variety of benefits were elicited through this training, such as performance, trust, and a faster learning rate for the autonomous agent [28]. A reproducible model for this type of human-autonomous agent team training has been developed and reviewed with these positive results and can be refactored to other contexts [29]. While training will vary widely from industry to industry, I4.0 provides a generalizable example for training with HATs known as "learning factories." These learning factories are meant to simulate future I4.0 factories and extend the skills currently held by employees [24]. The type of training should focus on cross-training if possible; however, merely exposing the user to the autonomous agent in a simulated work setting and making it clear how the autonomous agent is reaching its decisions should suffice. Training should also occur before any meaningful work is to be conducted, or if the user is working with an autonomous agent they have never worked with before.

As for human-AI interaction/teaming researchers, the primary point of training should be to familiarize the participant with the AI, giving them foundational transparency into how the autonomous agent makes decisions. This assertion applies to all research experiments that do not have training as a manipulated variable. The need to train participants (users) with the autonomous agent is necessary as past empirical work, and current industry operators have found it fundamental, making the application of any research results without training dubious.

Training is also heavily integrated with other core components of the framework, such as transparency, individual differences, and autonomy levels. Training can be utilized to complement individual differences by exposing human team members to well-designed autonomous agents, giving them a positive prior experience with an autonomous agent [12, 13], coupled with spatially based virtual training can give improved performance and trust in HATs [3]. Training is also a significant determining factor of transparency levels for the autonomous agent in HATs, as the training process familiarizes the human team member about its capabilities, functions, and tendencies. Finally, autonomous agents can benefit from proper training in these learning factories too [28]. The cross-training method enables autonomous agents to learn at rates faster and more efficient than those achievable using traditional learning models [28, 29]. Finally, the cross-training is a vital part of the framework, providing each member of the team with knowledge and expectations for each role. This training makes the "mental model" held by the autonomous agent much more accurate. This improvement

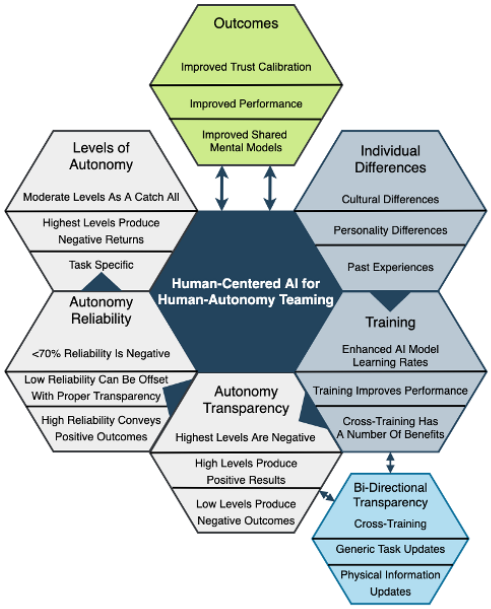

Figure 1: A conceptual framework for the design & implementation of autonomous agents in I4.0 applications

is because the autonomous agent's mental model becomes more global, containing knowledge about the entire team rather than its singular task. This training implementation is the driving force behind the bi-directional transparency concept in Figure 1, as training is capable of teaching the agent what actions and decisions to expect from the human and vice versa.

3.3 Bi-Directional Transparency

Bi-directional transparency is similar to bi-directional communication in HATs, as each concept shares the goal of communicating intent, current beliefs, goals, and potential obstacles [38]. The primary factor distinguishing the two concepts is bi-directional communications reliance on verbal or textual interaction, while the aforementioned bi-directional transparency does not rely on verbal or textual communication. Bi-directional transparency instead is implemented using various training methodologies, user interface design features, general inputs from the human, and general inputs from the autonomous agent. Bi-directional transparency is more applicable than bi-directional communication, given the current shortcomings of natural language processing (NLP) [16]. However, training methodologies like cross-training (see [28]), contribute to bi-directional transparency by allowing the agent to at least understand their team and potentially act in other roles if necessary.

The construct of bi-directional transparency asks that both the human and autonomous agent are transparent to the other. Transparency of the autonomous agent to the human should consist of the typical modalities like providing explanations of its actions and goals [23]. The autonomous agent should also be conveying consistent updates on intentions as any given situation or task progresses. Conversely, the human should provide similar information to the agent to enhance their transparency. This transparency could occur through physiological sensor readouts and generic inputs tied to context, given throughout the task execution. Psychophysiological sensor readouts have been suggested for use in adaptive automation in the past [2] and can be implemented here for enhancing transparency into the human decision-making process.

Such information would give autonomous agents enhanced visibility into things like stress levels and workload [2], which, when coupled with location information from GPS sensors, can give the autonomous agent information regarding where the human is in the task process and how well they are doing. Parasuramann and colleagues have already theorized such a feature in the form of adaptive automation. Adaptive automation features dynamically changing levels of independence and task load for the autonomous agent dependent on the human operator's psychophysiological state ([2]). The more stress and or workload the human operator is experiencing, the more ancillary work the autonomous agent would take, allowing the human operator to focus on their primary task. This psychophysiological information is then coupled with generic inputs from the human teammate that have the same operational goal as the updates from the agent about their respective decision-making process. These outputs come together to serve the function of explaining deviations from more expected actions based on the cross-training received by each team member. Providing these updates and information ensures the team mental model is updated rationally and with the necessary information behind it. Transparency is also highly beneficial for the human teammates, as outlined in Chen and colleague's 2018 work [6], and detailed further in the subsequent section on transparency.

Implementing bi-directional transparency for researchers in an experimental setting depends upon the unique circumstances of the experiment in question. For example, the human should have a transparent agent from prior training and interface elements, like those seen in Mercado and colleague's study, which implemented

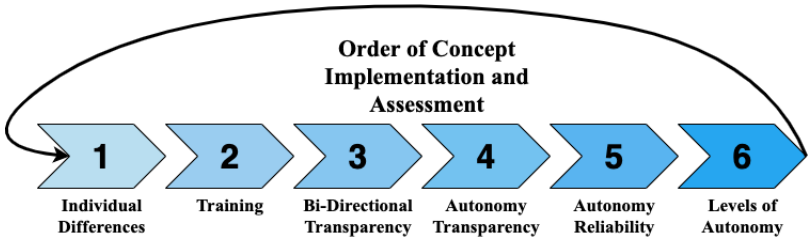

Figure 2: An order of the implementation & assessment of the relevant factors of human-autonomy teaming as outlined by the conceptual framework.

transparent icons to convey the certainty of the autonomous agent's decision making ([27]). The transparency for the agent will look much different depending on the experiment and could include things like the participant's current location (in a game), their heart rate and galvanic skin response (for stress) ([2]), virtually any information about the human teammate(s) that the autonomous agent may find useful in decision making. It should also be noted that for the human's transparency to take effect, the information must be included in the training cycle for the AI, allowing it to learn to use it in its decision-making process. This bi-directional transparency in experimental settings would be novel to the human-AI interaction/teaming literature in terms of its effects, as it has mostly focused on bi-directional communication up to this point [38]. However, results should show similar effects to bi-directional communication, which include enhancing the team's awareness, shared understanding, and making the system less brittle and more generalizable [38]. However, such effects could be gained without the effort and hassle of developing a true shared language between humans and AI with NLP but with physical information about the human, interface design decisions, and training.

3.4 Autonomy Transparency

Autonomous agent transparency is the knowledge the human in the human-automation team has about the agent, such as how accurate and informed the human is on the agent's ability, intent, decision-making process, and situational parameters [23]. Autonomy transparency extends this and encompasses what information an operator wants or needs under the various work and situational contexts [23]. Literature shows that transparency, in general, can produce positive results in outcomes such as performance, trust, trust calibration, perceived usability, and agreement [5, 27, 44]. Additionally, these favorable results to the human team members were yielded without penalty to their workload or response times [27]. That being said, transparency is still prone to induce detrimental effects on human team members in specific contexts. There is evidence that higher levels of transparency produce adverse outcomes on human team members, indicating potential diminishing returns or negative returns at extremely high levels [44].

HAT practitioners must take autonomous agent transparency seriously when implementing their HATs, as the construct interacts with many of the other constructs presented in the current model. Practitioners must engender appropriate levels of transparency in their HATs, which has much to do with the construct of training and how it is carried out. Training is a primary means to manipulate transparency, especially when coupled with user interface design features. The connection between training and transparency and its relationship with reliability, discussed in the subsequent section, makes transparency a critical factor affecting HAT outcomes. Thus, to ensure practitioners are reaping the positive benefits of transparency, the information overload that occurs at the highest levels of transparency must be avoided at all costs. Human teammates must know as much about the autonomous agent and its operation under a variety of contexts as relevant to their shared goal; anything more is just contributing to information overload and is unnecessary. Autonomous agent transparency is primarily created through the training and knowledge that human team members receive *about* their autonomous teammate and the subsequent hands-on training *with* the autonomous teammate. However, autonomous agents continuously learn and can change their behavior, meaning human teammates' mental models may become inaccurate over time. Accounting for changes in the autonomous agent's model can be done by implementing the bi-directional transparency seen in the framework (Figure 1). This bi-directional transparency is necessary as it gives insight into the decision-making process of the autonomous agent for the human, but the autonomous agent is also given insight into the human's decisions. This level of information sharing offers the benefits of transparency to the human and the agent, enhancing the team's ability to understand and operate with one another. This bi-directional transparency is brought on by various factors that include training, task-relevant inputs from the human, and task-relevant inputs from the autonomous agent, each covered in detail in its respective section.

Researchers in human-AI interaction/teaming should always ensure that transparency is installed into their experimental settings unless it is a variable of interest. At the minimum level, researchers should implement a training period described in the previous training subsection to ensure humans understand how to work with the agent and get an idea of how it's decisions are made. Bi-directional transparency should also be implemented if possible in the autonomous agent's training, offering simulated human inputs.

3.5 Autonomy Reliability

The matter of autonomous agent reliability is straightforward, as research has found, the more reliable the autonomous agent, the more positive every outcome examined will be [4, 10]. In this research, levels of reliability purely involved the error rate of the autonomous agent. A similar metric also allows for autonomous

agent transparency to be conveyed to the human teammate by displaying the agent's confidence in their decision. However, what makes autonomous agent reliability a critical facet to integrate within this framework is its interaction with autonomous agent transparency. Research has shown that autonomous agents with a reliability level below 70% produce such negative results that not implementing any autonomous agent is better [42]. However, if the autonomous agent's reliability is well known and communicated beforehand to the humans (i.e., proper transparency), the human team members will calibrate their trust accordingly. This calibration in trust then offsets these adverse outcomes to produce positive results despite the lower reliability of the autonomous agent [10, 27]. Thus, this effect allows practitioners to implement HATs within their facilities to make better decisions on whether or not to implement specific autonomous agents based on their reliability. If an autonomous agent has a reliability of 60%, but there are time and resources for proper training of its human teammates, then the benefits of that autonomous agent can still be realized. However, if there is not enough time or resources to train the human team members, the call can be made to keep that autonomous agent in a shadowing role until it can adapt its model and achieve a higher level of reliability. Finally, for research-oriented HATs, autonomy reliability should always be as high as possible unless it is a variable of interest.

3.6 Levels of Autonomy

Levels of autonomy refer to the degree to which an autonomous agent will make decisions and or take action with or without human input [31]. These autonomy levels, as operationalized by Parasuraman and colleagues, consist of ten different levels of automation [31]. In this scale, an autonomous agent operating at level 10 acts as a fully autonomous agent that makes each decision without human input and even ignores humans outright, while a level 1 agent is manual system operation by a human. Level 6 is where automation turns into autonomy, and systems begin to exert control over their own decisions, responding to changing contexts, and showcasing independent operation. Past research shows that a moderate level of autonomy produces the best balance in terms of performance and decreased situational awareness [45]. However, this scale has since been adapted by O'Neill and colleagues for a review of relevant human-AI teaming literature, which now states that levels 1 through 4 are no-autonomy/manual control, levels 5 and 6 are partial autonomy, and levels 7 through 10 constitute high autonomy. However, this modified scale still retains the verbiage describing each level, just as Parasuraman and colleagues did. For example, level 5 partial autonomy allows the computer to suggest a decision to the human, which the human will then accept or deny, while level 6 autonomy suggests a decision and will execute the decision unless the human vetos it in a pre-defined amount of time. High levels of autonomy begin with level 7 and are characterized by the agent making decisions without any human input at all, while no autonomy in level 4 and below are generally characterized by the agent merely making suggestions but never executing.

However, taking the time to choose a level of autonomy that allows the operator (dyad) and or teammates (triad or more) to maintain a sense of inclusion and engagement is of the utmost importance. Taking this extra time ensures the humans successfully remain in the loop and do not suffer from a lack of situational awareness, task/team complacency, and human error. Such adverse effects are typically seen at the highest levels of autonomy, indicating that these autonomy levels are not suited for any HAT scenario. Autonomous agents with such high levels of autonomy are designed in ways that inherently make them bad teammates; as previously mentioned, level 10 autonomy ignores human input outright. Applied HATs will employ a range of autonomy levels between these two extremes meaning autonomous agents working with humans must be designed with a level of autonomy that complements the context they operate within. For example, a HAT working in I4.0 predictive maintenance operating at level 7 (executes automatically and informs the human teammate) allows the human to remain in the loop, ensuring the human can keep their mental model of the facility accurate while reducing cognitive load compared to level 6 agent autonomy. This awareness enhances their ability to predict future faults by experience, enabling the human to utilize intuition to make suggestions back to the autonomous agent, improving the team as a whole. These various levels have significant differences between them and play a crucial role in implementing human-AI teams in practice or research settings. For example, level 7 autonomy informs the human of the decision it has made and executed, while level 8 informs the human only if they ask. There is a significant difference between these two, and the specific contexts where a variety of level 7 works well, a level 8 implementation may be entirely inappropriate, reduce performance, and possibly even dangerous.

3.7 Framework Interactions

As a framework, each component is reactive to other components, shown by the arrows in Figure 1, and briefly addressed in the discussion of each. As the framework is divided into three distinct components, the interactions will be outlined accordingly. The human-specific factors are individual differences and training, grouped as they focus on the human within the team and are up to the practitioner to design and implement. For example, individual differences can be leveraged by designing proper training techniques to engender more initial trust in the employees' eventual HATs, and this can be done by fostering a positive experience with autonomous agents during any training period. Training is another human-specific factor of the framework but is highly connected to autonomy transparency and bi-directional transparency, which serves as the bridge to the second domain of factors focusing on the autonomous agent.

The bi-directional transparency component acts as a bridge between these human-related and autonomous agent related factors. The training of both the human and the autonomous agent serves as a critical driver to bi-directional transparency, ensuring an understanding of each distinct role within the team. Accomplishing this training for the autonomous agent can be accomplished through the autonomous agent's model training, shown in Schelble and colleagues 2020 work [37]. In this implementation, an agent was trained to successfully perform all three roles of a team-based task, ensuring the autonomous agent knew what to expect from each

member of the team. However, to further improve upon this training, human team members must also be trained in as many roles as possible. Autonomous agent-specific factors, on the other hand, include the facets of autonomy transparency, reliability, and autonomy level. Autonomy reliability is then moderated by autonomy transparency, where an autonomous agent below 70% reliability can still be useful if the autonomy is appropriately transparent to the humans within the team. The level of autonomy that the autonomous agent takes on is then determined and informed by its reliability and transparency. Outcomes are then a result of the successful use of the framework and push a positive feedback loop, as proper transparency levels from training results in enhanced shared mental models that contribute to individual differences as past experiences.

4 CONTEXTUALIZING THE FRAMEWORK TO APPLIED HUMAN-AUTONOMY TEAMS

This paper identified the core factors that relate to HAT integration in applied settings. From this, a framework was created that was supported by the literature, its applications to applied HAT settings detailed, and interaction between factors outlined. This discussion then details how the framework can be leveraged in its entirety for use in applied HAT integration, how its utilization stands to increase trust in HATs, and how the framework is scalable based on the growth of applied HAT settings. This discussion allows the framework to be better understood in its usage and how it is forward-facing and implements scalability for various contexts throughout the growing HAT environment.

4.1 Example of Framework Usage for Enhancing Trust

The framework's actual benefits are seen from in-depth analysis and investment of effort and time into using or studying the human-AI team's interactions. When the framework's specific factors are considered carefully, such as the individual differences of the human teammates, their training with the autonomous agent, variable transparency based on autonomous agent reliability, and level of autonomy, careful planning and manipulation can take these various facets and produce a strong, well developed human-AI team. For example, one of the most basic goals for those implementing and or designing autonomous agents for HAT is trust, which is an excellent example factor for the presented framework to maximize. Starting with individual differences, positive prior experiences with autonomous systems lead to enhanced trust [12, 13]. These positive experiences with autonomy can be fostered during training [28], along with autonomy transparency and bi-directional transparency. This enhanced transparency engenders increased trust, enhanced trust calibration, and performance from the team [5, 27, 44], while also minimizing the effects of lower reliability autonomy [10, 27]. As an additional note, while proper training lends itself to increased trust and performance in teams that participate in cross-training, these effects are not limited to this single form of training, as seen in Cohen & Imada's 2005 work [7]. Putting all this together allows practitioners to design and implement their HATs properly. Here, the framework's benefits are highlighted for an applied HAT practitioner seeking to maximize trust in an already existing HAT or soon to exist HAT. The relationships between each of the factors are clearly defined and interact with one another to create a positive HAT assessment and implementation life cycle.

4.2 Scaling and Extending the Framework

This framework also can adapt and grow based on new advances in technology. For example, a potential future for autonomy is likely to include levels of autonomy that are much more specific than the standard scale used here. As autonomy advances into things like cars, smart homes, and applied facilities like medicine [40], this scale will likely be advanced or further adapted in the future. This framework will be enhanced by such changes as it can be extended to include general levels of autonomy and potential new and more specific levels of autonomy. This enhancement would accommodate small and medium enterprises with fewer resources for advanced autonomy in addition to encompassing any new and advanced levels of autonomy for larger enterprises. These adaptations are made by extending and enhancing the concepts within the framework, integrating new knowledge, and showcasing the framework's ability to act dynamically over time. This adaptability and scalability also applies to extremely context-specific applications of the framework, which could potentially be seen when applying it to highly dichotomous manufacturing facilities (vehicle production facility vs. silicon foundry). One facility may utilize mostly automation-based agents, while the other strictly utilizes autonomy-based agents. This difference introduces the potential to adapt the framework into a hybrid that takes on more aspects that specifically relate to human-automation teaming than the current focus on human-autonomy teaming.

5 CONCLUSION

The framework presented in this paper encompasses two decades of HAT research related to the interests of applied HAT practitioners. The framework informs those same practitioners on integrating HATs in applied settings, providing a meaningful theoretical contribution as HATs are on the cusp of being deployed across an innumerable number of settings around the world. With this framework, users of HATs can make informed decisions that have the potential to directly impact outcomes by ensuring HATs are trained, deployed, and integrated in such a fashion that enables both the humans and the autonomous agents to reach their full potential. As each facet of the framework is backed by specific prior HAT research, it can engender a variety of benefits. The framework is capable of targeting a variety of different end goals such as enhanced trust through proper training, accurate trust calibration via appropriate transparency, better performance based on accommodating prior experiences in the learning factories, or human teammates that remain in the loop because their artificial teammate utilizes the correct level of autonomy. As such, the framework allows practitioners of HATs to make informed decisions about their implementation that will make it smoother, more productive and accommodating.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1829008.

REFERENCES

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2016. TensorFlow: A System for Large-Scale Machine Learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. USENIX Association, Savannah, GA, 265–283. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi

[2] Evan A. Byrne and Raja Parasuraman. 1996. Psychophysiology and adaptive automation. *Biological psychology* 42, 3 (1996), 249–268. Publisher: Elsevier.

[3] Jessie YC Chen and Michael J. Barnes. 2010. Supervisory control of robots using RoboLeader. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 54. SAGE Publications Sage CA: Los Angeles, CA, 1483–1487. Issue: 19.

[4] Jessie YC Chen, Michael J. Barnes, and Michelle Harper-Sciarini. 2010. Supervisory control of multiple robots: Human-performance issues and user-interface design. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 41, 4 (2010), 435–454.

[5] Jessie YC Chen, Michael J. Barnes, Anthony R. Selkowitz, and Kimberly Stowers. 2016. Effects of agent transparency on human-autonomy teaming effectiveness. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 001838–001843. https://doi.org/10.1109/smc.2016.7844505

[6] Jessie YC Chen, Shan G Lakhmani, Kimberly Stowers, Anthony R Selkowitz, Julia L Wright, and Michael Barnes. 2018. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical issues in ergonomics science* 19, 3 (2018), 259–282.

[7] Joseph Cohen and Andrew Imada. 2005. Agent-based training of distributed command and control teams. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 49. SAGE Publications Sage CA: Los Angeles, CA, 2164–2168.

[8] Michael T. Cox. 2013. Goal-driven autonomy and question-based problem recognition. In *Second Annual Conference on Advances in Cognitive Systems 2013, Poster Collection*. Citeseer, 29–45.

[9] Selim Erol, Andreas Jäger, Philipp Hold, Karl Ott, and Wilfried Sihn. 2016. Tangible Industry 4.0: a scenario-based approach to learning for the future of production. *Procedia CiRp* 54, 1 (2016), 13–18.

[10] Xiaocong Fan, Sooyoung Oh, Michael McNeese, John Yen, Haydee Cuevas, Laura Strater, and Mica R. Endsley. 2008. The influence of agent reliability on trust in human-agent collaboration. In *Proceedings of the 15th European conference on Cognitive ergonomics: the ergonomics of cool interaction*. 1–8.

[11] Christopher Flathmann, Nathan McNeese, and Lorenzo Barberis Canonico. 2019. Using Human-Agent Teams to Purposefully Design Multi-Agent Systems. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 63. SAGE Publications Sage CA: Los Angeles, CA, 1425–1429.

[12] Feyza Merve Hafizoglu and Sandip Sen. 2018. The Effects of Past Experience on Trust in Repeated Human-Agent Teamwork. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 514–522.

[13] Feyza Merve Hafizoglu and Sandip Sen. 2018. Reputation Based Trust In Human-Agent Teamwork Without Explicit Coordination. In *Proceedings of the 6th International Conference on Human-Agent Interaction*. 238–245.

[14] Nader Hanna and Deborah Richards. 2015. The Impact of Virtual Agent Personality on a Shared Mental Model with Humans during Collaboration. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. 1777–1778.

[15] Hsiao-Ying Huang and Masooda Bashir. 2017. Users' trust in automation: a cultural perspective. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 282–289.

[16] Robin Jia and Percy Liang. 2017. Adversarial examples for evaluating reading comprehension systems. *arXiv preprint arXiv:1707.07328* (2017).

[17] Hyoung Seok Kang, Ju Yeon Lee, SangSu Choi, Hyun Kim, Jun Hee Park, Ji Yeon Son, Bo Hyun Kim, and Sang Do Noh. 2016. Smart manufacturing: Past research, present findings, and future directions. *International journal of precision engineering and manufacturing-green technology* 3, 1 (2016), 111–128.

[18] Kristin Lee. 2016. Artificial Intelligence, Automation, and the Economy. https://obamawhitehouse.archives.gov/blog/2016/12/20/artificial-intelligence-automation-and-economy

[19] Matti Krüger, Christiane B Wiebel, and Heiko Wersing. 2017. From tools towards cooperative assistants. In *Proceedings of the 5th International Conference on Human Agent Interaction*. 287–294.

[20] Erik Lagerstedt, Maria Riveiro, and Serge Thill. 2017. Agent Autonomy and Locus of Responsibility for Team Situation Awareness. In *Proceedings of the 5th International Conference on Human Agent Interaction*. 261–269.

[21] Jay Lee, Behrad Bagheri, and Hung-An Kao. 2015. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing letters* 3 (2015), 18–23.

[22] Jay Lee, Hossein Davari, Jaskaran Singh, and Vibhor Pandhare. 2018. Industrial Artificial Intelligence for industry 4.0-based manufacturing systems. *Manufacturing letters* 18 (2018), 20–23. Publisher: Elsevier.

[23] Joseph B. Lyons. 2013. Being transparent about transparency: A model for human-robot interaction. In *2013 AAAI Spring Symposium Series*.

[24] Mohammed M. Mabkhot, Abdulrahman M. Al-Ahmari, Bashir Salah, and Hisham Alkhalefah. 2018. Requirements of the smart factory system: a survey and perspective. *Machines* 6, 2 (2018), 23.

[25] Deborah J MacInnis. 2011. A framework for conceptual contributions in marketing. *Journal of Marketing* 75, 4 (2011), 136–154.

[26] Nathan J McNeese, Mustafa Demir, Nancy J Cooke, and Christopher Myers. 2018. Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human factors* 60, 2 (2018), 262–273.

[27] Joseph E. Mercado, Michael A. Rupp, Jessie YC Chen, Michael J. Barnes, Daniel Barber, and Katelyn Procci. 2016. Intelligent agent transparency in human–agent teaming for Multi-UxV management. *Human factors* 58, 3 (2016), 401–415.

[28] Stefanos Nikolaidis, Przemyslaw Lasota, Ramya Ramakrishnan, and Julie Shah. 2015. Improved human–robot team performance through cross-training, an approach inspired by human team training practices. *The International Journal of Robotics Research* 34, 14 (2015), 1711–1730.

[29] Stefanos Nikolaidis and Julie Shah. 2013. Human-robot cross-training: computational formulation, modeling and evaluation of a human team training strategy. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 33–40.

[30] Thomas A. O'Neill, Nathan J. McNeese, Amy Barron, and Beau G. Schelble. 2020. Human-Autonomy Teaming: A Review of the Empirical Literature. *Manuscript submitted for publication* (2020).

[31] Raja Parasuraman, Thomas B. Sheridan, and Christopher D. Wickens. 2000. A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans* 30, 3 (2000), 286–297.

[32] H. S. Park. 2013. From Automation To Autonomy-A New Trend For Smart Manufacturing. *DAAAM international scientific book* (2013).

[33] Cristina Renzi, Francesco Leali, Marco Cavazzuti, and Angelo Oreste Andrisano. 2014. A review on artificial intelligence applications to the optimal design of dedicated and reconfigurable manufacturing systems. *The International Journal of Advanced Manufacturing Technology* 72, 1-4 (2014), 403–418. Publisher: Springer.

[34] Eduardo Salas, Nancy J. Cooke, and Michael A. Rosen. 2008. On teams, teamwork, and team performance: Discoveries and developments. *Human factors* 50, 3 (2008), 540–547. https://doi.org/10.1518/001872008x288457

[35] Eduardo Salas, Terry L. Dickinson, Sharolyn A. Converse, and Scott I. Tannenbaum. 1992. Toward an understanding of team performance and training. (1992).

[36] Michael Schaarschmidt, Alexander Kuhnle, and Kai Fricke. 2017. TensorForce: A TensorFlow library for applied reinforcement learning. *Web page* (2017).

[37] Beau G. Schelble, Lorenzo Barberis Canonico, Nathan McNeese, Jack Carroll, and Casey Hird. in press. Designing Human-Autonomy Teaming Experiments Through Reinforcement Learning. In *Proceedings of the Human Factors and Ergonomics Society 2020 Annual Meeting*. SAGE Publications Sage CA: Los Angeles, CA, Virtual.

[38] R. Jay Shively, Joel Lachter, Summer L. Brandt, Michael Matessa, Vernol Battiste, and Walter W. Johnson. 2017. Why human-autonomy teaming?. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 3–11. https://doi.org/10.1007/978-3-319-60642-2_1

[39] Thomas Z. Strybel, Jillian Keeler, Natassia Mattoon, Armando Alvarez, Vanui Barakezyan, Edward Barraza, James Park, Kim-Phuong L. Vu, and Vernol Battiste. 2017. Measuring the effectiveness of human autonomy teaming. In *International Conference on Applied Human Factors and Ergonomics*. Springer, 23–33.

[40] Dayong Wang, Aditya Khosla, Rishab Gargeya, Humayun Irshad, and Andrew H. Beck. 2016. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718* (2016).

[41] David A Whetten. 1989. What constitutes a theoretical contribution? *Academy of management review* 14, 4 (1989), 490–495.

[42] Christopher D. Wickens and Stephen R. Dixon. 2007. The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science* 8, 3 (2007), 201–212.

[43] H. James Wilson and Paul R. Daugherty. 2018. Collaborative intelligence: humans and AI are joining forces. *Harvard Business Review* 96, 4 (2018), 114–123.

[44] Julia L. Wright, Jessie YC Chen, Michael J. Barnes, and Peter A. Hancock. 2016. The effect of agent reasoning transparency on automation bias: an analysis of response performance. In *International Conference on Virtual, Augmented and Mixed Reality*. Springer, 465–477.

[45] Julia L. Wright, Jessie Y. Chen, Stephanie A. Quinn, and Michael J. Barnes. 2013. *The effects of level of autonomy on human-agent teaming for multi-robot control and local security maintenance*. Technical Report. Army Research Lab Aberdeen Proving Ground, MD.