IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received March 30, 2021, accepted April 27, 2021, date of publication May 7, 2021, date of current version May 18, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3078298

Systemic Oversimplification Limits the Potential for Human-AI Partnership

JASON S. METCALFE, BRANDON S. PERELMAN, DAVID L. BOOTHE, AND KALEB MCDOWELL, (Senior Member, IEEE)
U.S. Combat Capabilities Development Command Army Research Laboratory, Aberdeen Proving Ground, MD 21005, USA

Corresponding author: Jason S. Metcalfe (jason.s.metcalfe2.civ@mail.mil)

This work was supported by the U.S. Department of Defense through the Human Autonomy Teaming Essential Research Program of the U.S. DEVCOM Army Research Laboratory.

ABSTRACT The modern world is evolving rapidly, especially with respect to the development and proliferation of increasingly intelligent, artificial intelligence (AI) and AI-related technologies. Nevertheless, in many ways, what this class of technologies has offered as return on investment remains less impressive than what has been promised. In the present paper, we argue that the continued failure to realize the potential in modern AI and AI-related technologies is largely attributable to the oversimplified, yet pervasive ways that our global society treats the relationship between these technologies and humans. Oversimplified concepts, once conveyed, tend to perpetuate myths that in turn limit the impact of such technologies in human society. To counter these oversimplifications, we offer a theoretical construct, which we call the landscape of human-AI partnership. This construct characterizes individual capability for real-world task performance as a dynamic function of information certainty, available time to respond, and task complexity. With this, our goal is to encourage more nuanced discourse about novel ways to solve challenges to modern and future sociotechnical societies, but without defaulting to notions that remain rooted in today's technologies-as-tools ways of thinking. The core of our argument is that society at large must recognize that intelligent technologies are evolving well beyond being mere tools for human use and are instead becoming capable of operating as interdependent teammates. This means that *how* we think about interactions between humans and AI needs to go beyond a ''Human–or–AI'' conversation about task assignments to more contextualized ''Human–and–AI'' way of thinking about how best to capitalize on the strengths hidden within emergent capabilities of unique human-AI partnerships that have yet to be fully realized.

INDEX TERMS Human-AI partnership, human-autonomy teaming, sociotechnical systems, AI ecosystems, function allocation, task complexity, capability, use cases, implementation.

I. INTRODUCTION

Our global society is in the midst of what some consider to be one of the most sweeping and disruptive periods of technology evolution in history [1]–[3]. The exponential growth in the sector of artificial intelligence (AI) and AI-related technologies[1] – here, defined as the domain concerned with intelligent agents that have sensing, perceiving, rudimentary reasoning, and/or learning capabilities – is providing unprecedented opportunity for advancement of human society. However, caution is warranted as the speed of development and high-end potential of these technologies is not immutable. The ''AI winter'' that devastated the field in the 1970's [4]–[6] serves as a persistent reminder that forecasts of AI, from promises to threats, are susceptible to exaggeration. Scholars from a breadth of backgrounds have expressed concern about the oft-neglected limitations in state of the art approaches, many of which will only be resolved by new discoveries (c.f. [6]). Here, we argue that the limitations of current technology integration approaches arise from oversimplified assumptions about the human-AI relationship. We then argue that these oversimplifications may be

The associate editor coordinating the review of this manuscript and approving it for publication was Zhiwei Gao.

[1] This paper is about a general class of technology that is diverse and lacks a broadly accepted descriptor. Our intent is to communicate effectively, and because we lack an acceptable general term, we compress the phrase ''AI and AI-related technologies'' into ''AI'' through most of our discourse. This, of course, conflates expert systems, machine learning, autonomy, robotics, intelligent sensing (and many others) with ''AI''. The irony in our simplification is not lost; however dubious, here we prioritize readability over precision in this limited way.

overcome by developing intelligent sociotechnical ecosystems, which would be built upon multiple, coexisting interaction dynamics that support complex human-AI partnership; in more direct terms, by enabling effective teamwork.

As AI and AI-related technologies are integrated into our lives, the methods for introducing them, as well as for effectively integrating them with our society, have been the source of involved, long-term debate. Such debates can be found at least as early as the 1930's [7], with specific discussions about designing effective ''man-machine systems'' [8] published the same year that also welcomed the first robot into existence [9]. Since then, varied frameworks have been promoted for systematically defining roles and responsibilities in human-machine interactions [10]–[19], indicating that there is still no established and broadly accepted *correct* method for achieving integration in all situations. Part of the reason for this apparent lack of consensus, we argue, is that the core research community is exceptionally diverse. Whether hardware specialists, algorithm developers, human factors engineers, creative designers, business professionals, technology transition managers, marketing specialists, or end-users, most parts of society are playing a role in this exciting global transformation. While the combined expertise of multidisciplinary teams is needed, the diverse and specialized perspectives that individuals bring to the table can make it difficult to form a shared vision [20]. In these highly multidisciplinary ventures, limitations on language and differences in training and priorities can systematically perpetuate oversimplifications that are then commonly shared among stakeholders across technical literacy barriers [21].

In the present paper we offer our perspective on this circumstance by first establishing a backdrop of three oversimplified assumptions about the human-AI relationship, and we go on to offer a theoretical construct that we have developed to facilitate collaborative discourse about strategies to integrate this class of technologies into new intelligent sociotechnical ecosystems with humans. The ultimate aim is to expand the global discussion around how we, as a society, may form effective human-AI partnerships. We have selected a few oversimplifications that, like others, seem to arise from viewing human-AI interactions as monolithic, using a single construct for understanding division of labor. Rather, we argue that the fundamental nature of future human-AI partnerships is task relative, depending on the certainty of information forming the basis of the problem, the amount of time available to resolve the situation, and, most critically (and most often ignored), task complexity.

As AIs become more advanced, our patterns of interaction with them may be expected to progress as well. Moving our vision away from the more clearly differentiated, and simpler, roles that exist between craftsman and tool, we anticipate that far more complex and interdependent dynamics will emerge as humans and AIs are brought into intelligent ecosystems enabled by varied team-like partnerships and interaction dynamics [22], [23]. Viewing interactions between AI and humans in the context of certainty, time, and complexity should clearly emphasize the reality that there is not one sort of interaction that must be considered and supported. Rather, as with large social systems, there are myriad ways that humans and AIs may cooperate or compete. Therefore, no single paradigm will appropriately address the question of how best to integrate the two. The growing community of interest around human-AI teaming must, we argue, come to agreement on how to address the problem space, as well as how we convey important understandings about it; here, it is our aim to offer the theoretical construct that we call the *landscape of human-AI partnership* to support these goals.

II. (SOME) OVERSIMPLIFICATIONS IN CONSIDERING AI

One does not need to look very hard for examples of science and technology inspiring thoughts that, while creative and innovative, are also divorced from our physical reality. It has been said, for instance, that Mary Shelley's concept of *Frankenstein*, written in 1818, was at least partially inspired by her understanding of Dr. Humphry Davy's writings and public demonstrations involving animating cadavers with then, newly discovered electricity [24]. More related, *I, Robot*, which depicted a society capable of producing an AI that could emulate and replace humans, was first published by biochemistry professor and sci-fi author Isaac Asimov in 1950, four years ahead of the very existence any robot [25]. Though these examples may seem a bit tangential, their relevance is that they provide context for our main argument that humans tend to develop mental concepts of science and technology that oversimplify and inaccurately represent (or obfuscate) the underlying natural principles (e.g. Asimov's ''three rules'' as a sufficient ethical framework, or Shelley's idea of reanimation with lightning).

Here, we note that there are many ways in which we believe that the general discourse about AI within society is oversimplified; yet, the present work is not meant to be a complete treatise on them. Our specific focus is on oversimplifications that most closely relate to human-AI partnership. The present section frames our argument in a small set of oversimplifications that we have repeatedly encountered in our collective experience within the research community; we have not conducted any surveys or formal inquiries to determine these, rather we use them to illustrate common themes. While the oversimplifications themselves are expository, we support them with many examples from both technical and lay writing. Importantly, we believe that this integrative, cross-disciplinary discussion is important, because the oversimplifications constitute barriers to accurately understanding the challenges of realizing future human-AI partnerships. We follow this in the latter portions of this paper by articulating our perspective as a theoretical construct that is needed to replace old and regressive notions with those that are more suitable to the advancing technical capabilities of our modern world.

Oversimplification 1: ''AI will make humans obsolete.'' This oversimplification reflects the belief that AI is inherently superior, will necessarily learn to outperform humans, and therefore the human will become redundant. In fact, humans have been debating their own replacement by machinery of various types for as long as, if not longer than, machines and specialized instruments have been around [7], [25]; to wit, the very word 'robot' derives from a play by Karel Čapek that first debuted in 1921, over 30 years before the first robot existed (and the topic of the play was a robot uprising against humans). This oversimplification is not particularly specific to AI; Keynes coined the term "technological unemployment" to describe periods of human obsolescence produced by a mismatch between labor availability (in type and quantity) and needs [26] (see McGaughey [27] for more detailed discussion). When taken to the logical limit, this implies that as AI becomes more pervasive and generally capable, human action will move into obsolescence.

Today, beyond the kinds of questionable implications discussed above, the nature and timing of this obsolescence, and the details of precisely which jobs are vulnerable, are debated in very real ways. Some experts continue to express the belief that AI will surpass humans at all tasks within this century, specifically by 2060 (and some say earlier), while others do not foresee this happening for more than 200 years (and some say never) [28]. Generally, jobs traditionally requiring a "human touch" are frequently discussed as immune to replacement by AI (see [29] further reading). Nevertheless, we see that across sectors, when AI is introduced, one of the first fears to arise is that humans in that sector will be replaced. Further, non-experts tend to expect these replacements will be more widespread and occur sooner than experts [30]. In general, fear of human replacement appears to have originated from the intersection of notions that AI has an inherent and unequivocal advantage and, moreover, that societal needs will be largely limited to the domain where the putative AI advantage dominates all other considerations. At its root, this oversimplification misses that functional, real-world tasks are not as simple as depicted. That is, because a task seems simple for humans, it is often wrongly assumed that it must also be simple to automate; an easy example here is to explore the vast research literature attempting to explain even the simplest human behaviors – like upright standing [31] or rhythmic movements of a single finger [32]; complexity exists even for the simplest of biological systems. The oversimplified notion that humans will be made obsolete also misses that there are viable strategies to offset the differences in capabilities. Here, we provide two examples of how this oversimplification frequently manifests:

- Hardware processing speeds allow much faster information handling than humans, and certain events will occur so fast that humans cannot possibly be "in-the-loop". In a particularly morbid example, this oversimplification plays out as an ethical, life and death situation unfolding too rapidly for human intervention: a self-driving car with only milliseconds to react must decide which of the vehicle's occupants, pedestrians crossing the street, or their pets, will perish [33]. More nuanced examples include contemporary military thinking that AI-enabled warfare will be separable from "human spaces" [34], and perspectives on the human's role in domains like high-speed trading [35], wherein real consequences have manifested from run-away AI (e.g., the 2010 'Flash Crash' [36]). Logically, this suggests that humans should only be an observer, supervisor, or end-user – at least until such time as humans consent to surgical brain implants to improve their bandwidth for communicating on par with AIs [37], [38]. That is, when occupying a critical role in a control process, humans are usually expected to be slow and error prone, and thus the natural tendency is to want to remove them from time-critical paths. Yet, the general belief that human mental processing is inherently slower than AI has also been recently challenged. While in 2015 it was already trivial to show computers besting people for simple computations, two doctoral students at UC Berkely and Carnegie Mellon remained unconvinced, believing that a fair test of processing speed must be more functional than simply comparing the smallest possible unit of computation (e.g. a single floating point operation). To enable a fair comparison between brains and processors, the students devised a method to quantify how quickly an information network can be navigated and searched [39]. Using this functional measure (called TEPS, or Transversed-Edges Per Second), the investigators concluded that the human brain could at least match a supercomputer and is likely faster by a factor that approaches 30 times [40]. Even in the face of a significant speed differential, replacement of human elements is not the only logical option for human-AI partnership; that is, there are potentially helpful mechanisms that a mindset of human replacement neglects, such as pre- or mid-task criterion changes to preemptively adjust decision thresholds (e.g., to improve signal detection performance [41]), unsupervised algorithms that do not require, but nevertheless can accept and integrate human guidance for net improvement [42], and techniques that use AI to inform human decision making and other potentially more advanced techniques, including simulation-based, online, or even 'faster than real-time' decision-support systems [43], [44], which may ultimately allow for humans to influence, prepare, and tune the system for tasks that otherwise happen at super-human speeds.
- Broadening capabilities, combined with inherent storage, access, and processing capacity, will allow AI to encroach on high-level decision-making roles currently occupied by humans (see [45]–[47], and related ethical discussion in [48]); as such, some naturally expect that AI will also displace humans from roles like management [49] and human resources [50]. This oversimplification fails to consider the very real bounds on the types of problems that AI can solve, and how well it can solve those problems. Some of the earliest AI research was focused on developing deterministic

solutions to highly generalized problems, such as finding the shortest path through a graph [51], [52], optimally or near-optimally, in polynomial time. However, the limitations of such deterministic, analytic approaches were recognized early on; many real-world problems are simply too computationally complex to be solved in this way [53]. Today, real-world problems are generally approached through approximation algorithms, which are themselves bounded by computational requirements. For instance, five of the primary application areas for deep learning, including image classification and object detection, have been discussed as limited in this way [54]. Further, many real world problems involve factors like uncertainty, moral and ethical ambiguity, and so-called common sense – combined, these factors produce situations in which all solutions reflect trade offs that may be appropriate, and choosing one may depend on contextual factors that the AI may not be trained to access or understand. Mechanisms ranging from hierarchical rules, democratic voting, and other forms of joint decision-making and learning may provide options for developing human-AI partnerships that support complex behavior in ways that are superior to either entity acting independently [55].

Oversimplification 2: "Human intelligence is unique and irreplaceable by AI." This is essentially the contrapositive of the first oversimplification, and therefore has also been around for a long time. For instance, in a 1935 entry in the *Journal of Philosophy*, Kantor argued this point using the example of physicians who were increasingly reliant on their instruments (rather than their minds) to make diagnoses and conclusions [7]. Kantor argued, "*The thinker is prior to the machine. Machines or formulae can only help in our study; they cannot initiate or direct an investigation. Only the thinker can do that.* (p. 378)". One way that this oversimplification frequently manifests is in the form of lists (e.g., [56]) or even academic models (e.g., [57]) of the types of jobs that are believed to be safe from AI-induced technological obsolescence. This belief also arises from general observations in human sciences that have provided insights into the often hidden power of the brain (see breakout box "Human Intelligence – AI's Super Power").

HUMAN INTELLIGENCE – AI'S SUPER POWER

General intelligence in the academic field of artificial intelligence refers to a system with a range of human cognitive capabilities; effectively an attempt to simulate the human mental behavior (c.f. [58]). Importantly, in human science fields such as psychology and neuroscience, the concept of general intelligence is not well specified or even commonly defined [59]. This lack of agreement and understanding is a byproduct of the complexity of the human brain. Weighing in at about 1.4 kg and containing a mass of 100,000,000,000 nerve cells (not counting all the critical support cells) organized into myriad specialized architectures, the brain generates behaviors through numerous interconnected and often intertwined adaptive networks that produce complex, dynamically emergent activity [60].

While humans possess a unique intelligence,
the mechanisms of that intelligence are
incompletely understood.

Mental experience does not solely arise from the brain either. This complex structure does not produce the unified notion of "general intelligence" in humans, at least not on its own. Rather human intelligence is the product of a collection of capabilities that are traditionally considered "cognitive" (e.g., quantitative reasoning, fluid reasoning, visual-spatial processing, knowledge, working memory) and "non-cognitive" (e.g., empathy, interpersonal skills, emotional maturity). Further, human intelligence is considered a species-wide trait; yet the manifestations of each cognitive and non-cognitive capability vary widely between and within individuals.

Human intelligence is not monolithic,
but instead underlies a collection of evolutionarily
critical core attributes.

While its mechanisms are incompletely understood, the core attributes that arise from human intelligence are extremely valuable for sustaining functionality and capability in dynamic, adaptive, complex environments. Human survival has directly depended on effective adaptability, creativity, common sense, forethought, heterogeneous approaches to decision making, and leadership. Other attributes have their role in survival as well including humor, integrity, moral reasoning, emotional expression, and storytelling.

As our sociotechnical society continues to evolve, some steadfastly hold the human brain as unique and not fully replicable in either form or function, while others continue to argue that AI will outpace human intelligence. We argue that, while the underlying mechanisms are not fully understood, human intelligence has uniquely evolved and thus bears distinct strengths and weaknesses relative to AI. This intelligence, if effectively partnered with AI, will be a superpower; creating effective, adaptive, moral human-technology unions that outpace and outlast other forms of technology.

This oversimplification likely stems from the general appearance of certain human qualities as being scientifically intractable, such as the so-called "hard problem of consciousness" [61], which some argue are simply inaccessible to the human mind. Logically, as it is argued, if there are things *about humans* that are fundamentally inaccessible to understanding *by humans*, then it is also unlikely that these things could be accurately or precisely modeled by any human efforts. Without adequate models, insights into novel and innovative strategies for human-AI partnership will also remain limited. Beliefs about the putative incompatibility of AI with certain "soft" tasks are widely held by experts, even experts in AI. For example, Kai-Fu Lee, CEO of innovation Ventures and former vice president at three major US Silicon Valley-based companies, stated that jobs requiring uniquely-human attributes are impervious to AI-induced obsolescence, including complex and strategic jobs, and jobs requiring creativity or empathy [56]. This Human–or–AI perspective, as also manifested in the first oversimplification, misses the potential for mutually beneficial and synergistic operation. Two examples are illustrated below:

- Human intelligence is so unique that AI will fail to ever achieve its attributes, e.g., human-like ethical behavior, moral reasoning, and common sense. We subscribe to the notion that human intelligence, though heavily researched, remains incompletely understood. Academic understandings about human cognition have largely not been articulated in ways that readily translate into rules or logical structures that may be implemented in computational, cognitive systems.

However, this does not make said attributes inaccessible to AI in any absolute sense. For example, recent advances have allowed machines to detect emotion from speech [62] and facial expressions [63]. Through a combination of emerging technologies and cultural shifts in human expectations for interactions with machines, we may see AI being increasingly deployed against even ''soft'' human-oriented aspects of tasks; examples include, triage of a large number of inputs to present to humans, as in identifying and prioritizing highly-distressed callers to a support hotline, or service industry tasks that have, until now, been considered part of the human-only domain [64]. It is also reasonable to expect that breakthroughs in these human sciences may translate to downstream breakthroughs in AI capability. We anticipate that over sufficient time, AI-enabled systems will have the capacity to exhibit complex but perhaps qualitatively different high-level cognition sufficient to support those jobs currently believed to be in the ''only human'' domain.

- Many continue to believe that humans possess an exclusive intelligence, for example the power of creativity, that permits them to complete tasks that are inaccessible to AI. This presumes that humans are both necessary and sufficient to complete tasks requiring cognitive functions like creative ideation, and that jobs requiring these functions are not susceptible to technological unemployment [56]. Advances in AI aimed at creativity have challenged this perception (c.f., AI that can produce artistic images after training on a data set comprising 5 centuries of Western paintings [65]). Similarly, looking back to Kantor's argument for human scientists as the generators of ideas, we note how AI is even currently being developed to support semi-automated hypothesis generation [66], and to produce other forms of novelty like creating unique digits starting with a basis set of existing digits [67]. Perhaps the greatest gains will be realized when such human attributes are augmented by the rapid processing capabilities of AI. Mechanisms that merge AI and human intelligence, such as interactive machine learning approaches (e.g., learning from demonstration generally [68] as well as more recent hybrid methods [42]), can enable rapid AI adaptation by enabling non-expert users to train and retrain the agent as needed. Collaborative design paradigms, in which AI rapidly generates outputs based on human design specifications, empower human-AI teams to improve their performance in objectively-measured engineering tasks, like designing better quadrotors [69], as well as in more subjective artistic tasks like fashion clothing design [70]. Unique approaches to breaking down problems, like those found in the Human Computation and ''gaming with a purpose'' literature (as first described by von Ahn [71]), can be used to identify significant roles for AIs in these environments. Rather than humans being required to perform specific roles and tasks alone, human-AI partnerships will allow progress towards ''super-intelligent'' teams that enhance processes and improve overall performance [55].

Oversimplification 3: ''Integrating AI is as easy as assigning tasks based on individual strengths and weaknesses.'' This key oversimplification originated from the work of Paul M. Fitts in the 1950s [72], and remains widely-held today; that is, humans and AIs uniquely excel in qualitatively different, mutually exclusive functional domains. In the human factors literature, this has been discussed as a generalized ''HABA-MABA'' (humans-are-better-at, machines-are-better-at) perspective, which encourages use of substitution-based function allocation methods [12], [73], [74]. Viewing the world through this lens leads to stereotyped beliefs about capability differences (e.g., humans are slow but flexible; AIs are fast and precise, but rigid) that inform the design of simplified human-AI function allocation schemes, wherein tasks are assigned exclusively to one agent type or the other. On this basis, we will refer to HABA-MABA and similar concepts as belonging to a generalized ''Human–or–AI'' perspective through the rest of this paper.

In manufacturing, the Human–or–AI perspective, and its resulting reliance on substitution-based function allocation methods, enjoyed early success due to the segmented nature of the work (i.e., tasks that alternately require flexibility versus speed and precision) and the need to physically separate human and robot workspaces for safety [75]. While this perspective logically extends to other fields in which the interaction between human and agent is physical, it may not be as applicable when the interaction is more cognitive in nature. Evolving hybrid architectures, which combine bottom-up processing (for example, by neural networks) with top-down symbolic representations, challenge this persistent Human–or–AI perspective's hard boundaries by allowing machines to complete a wider variety of tasks, including those that are 'cognitive' [76]. This view quite possibly represents *the* archetypal oversimplification within the domain of human-AI partnership. We believe this view ignores a vast middle ground – the ''gray areas'' at the soft-boundaries between human and AI excellence. Here, we discuss the difficulties in characterizing human and AI capability sets respectively in order to demonstrate the folly in assuming that a simple Human–or–AI function allocation will provide general solutions that are well-suited (or even useful) across contexts and circumstances in the real-world.

- Human–or–AI framing oversimplifies the fact that human decision making is not bound to particular time scales or levels of accuracy; that is, it is neither always fast or always slow, nor is it reliably accurate or predictably error-prone across contexts [77]. Humans have evolved biological mechanisms and developed psychological strategies that they can deploy to solve complex problems and make difficult decisions with extreme efficiency, even in the absence of complete certainty or time to formulate a complete response. Humans accomplish this by reducing the dimensionality of the problem, reformatting it such that it is more readily consumable (for example, translating a computational problem to the visual modality), or by ignoring part of the information [78]. While often effective, the same characteristics that make these biological and psychological mechanisms rapid and adaptive can also manifest as maladaptive biases. Decision heuristics may impair peoples' ability to accurately judge event probabilities [79] and, likewise, people can get "locked-in" to particular solutions, which blinds them to alternatives (i.e., confirmation bias [80]). Moreover, humans have a tendency misinterpret their own capabilities relative to the capabilities of others [81], believing themselves to be better-than-average [82]. Therefore, we propose that such function allocation methods will continue to neglect critical information about human decision making: it is not slow compared to AI, per se, but rather is geared toward generating actionable (if biased) solutions within the biologically-relevant constraints of the human brain, and there are conditions where this is useful.

- Similarly, while AI successes are broadly disseminated in the public domain, the limitations tend to only be well-understood in the computer and computational sciences; this understanding has yet to become common in sciences that are adjacent to AI, much less in the broader public domain. The news that an AI algorithm has beaten an expert human in a particular game can, and often does, promote the perception that the AI is more intelligent than the human counterpart (at least, at that particular task). However, most experts understand that the truth is more nuanced. Deep learning models, for example, have gained considerable fame in recent years due to their ability to process images extremely quickly and with high confidence, or to generate new video using trained encoders (e.g., "deepfakes"; [83]). However, within the past decade, we have also learned that deep learners fail unpredictably, because they use complex data features to make classifications in ways that differ from humans (that is, the strongest predictors for an AI are not necessarily the features that are most salient to a human [84]). These classifiers are highly susceptible to adversarial attack [85] and often misclassify objects based on manipulations as simple as adding a border to the image [86] or rotating the object slightly [87]. Likewise, deep learners have been shown to make misclassifications of a sort that a human never would, such as classifying abstract black and white pattern images as an assault rifle, or even more nefarious, with very simple manipulations like adding a small icon to the image, turning a stop sign into a speed limit 45 sign [88]. That is, owing to their continued reliance on training data sets, which are unlikely be fully exhaustive, even well-trained AIs can be "fooled" to produce catastrophic outputs upon which higher-level decisions are then made (e.g. a self-driving car suddenly accelerating rather than stopping).

Indeed, there are myriad more ways in which the human-AI relationship is oversimplified and misunderstood within various elements of our modern society. The three oversimplifications presented here were chosen based on their recurrence as barriers to communicating and achieving common understandings of the challenges and opportunities for integrating AI into humans' daily lives. Frequently, it seems that discussions of AI relative to humans end up devolving into debates fueled by the exact notions captured above: "Human jobs will be lost to AI and robots." "AI can't possibly replace artists, therapists, and managers." "Complex decisions will be offloaded to advanced, intelligent computers and yet, we will always need human mediators when other people do not accept those decisions." Though new AI and machine learning approaches are continually being developed and brought to bear within each of these exemplar sectors, few have seen marked success in terms of broad acceptance and full integration as a *de facto* part of our society.

For the rest of this paper, we seek to persuade our readers that societal and individual thinking must go beyond and even challenge such oversimplified assumptions. We begin with a brief discussion of our understanding of the roots of the Human–or–AI ("HABA-MABA") perspective, as well as why we believe this logic, while valid, only holds within specific limits. More importantly, we argue that a broad re-conceptualization is needed to support effective human-AI partnerships, particularly in complex environments where the Human–or–AI logic breaks down. Finally, we explore some of the opportunities that may evolve by displacing the view of technologies as tools, or task-specific surrogates for humans, by discussing use-cases wherein a broad variety of human-AI partnerships may manifest in an intelligent sociotechnical ecosystem, several of which already exist.

III. ESTABLISHING A COMMON PERSPECTIVE

A major challenge that we see for advancing effective human-AI partnerships, is the reality that truly transformational progress requires diverse input from multidisciplinary teams of experts drawn from very different domains. Within the cross-disciplinary space that exists at their intersection, lies the challenge of communicating across very domain-centric, specialized lexical boundaries. Directly, we believe that a major factor underlying oversimplified assumptions about AI is an inherent difficulty in communicating and working across these domain-centric boundaries. More importantly, as scientific concepts are explored and developed in more depth, surface knowledge becomes less useful for understanding the capabilities and limitations of applications of those concepts. Recognizing our need to collaborate in these complex problem spaces, an important motivation for the present paper is to level-set and offer some concepts that may support the development of common goals and strategies. In this section, we specifically seek to set aside the oversimplified ways of thinking and begin to establish a more technically accurate perspective by exploring the more general nature of the shared problem space for human-AI partnership. Substantively, we offer a theoretical construct, which we call the *landscape of human-AI partnership*, with the intention of progressing towards more unified and domain-general ways of thinking and talking about building robust and effective human-AI partnerships.

A. AN EVOLVING CONCEPTUAL FRAMEWORK

As the background provided above illustrates, an oversimplified Human–or–AI perspective underlies many discussions about role definitions for humans and AI. This, often tacitly accepted, "either-or" perspective tends to drive questions towards function allocation-based solutions. "*Why not just assign tasks to the human that they do best and let the AI do the rest? Why can't a human just supervise as long as they are provided the tools they need to fix problems that the AI can't handle?*" These kinds of questions are common and reflect a Human–or–AI perspective. Further, such a focus appears to assume that the task of delineating human and AI capabilities in order to enable efficient function allocation is (or will be) relatively easy. We suggest that such concepts can only be applied both easily and effectively for simple tasks – well-structured tasks in which the goals are clear, the actions needed to achieve them are well-defined, and the response can be expected to occur as intended. We also suggest that these concepts will otherwise be quick to fail as tasks become more complex. Simple tasks are often invoked as example use cases for successful function allocation, but all the while ignoring the likely occurrence of suboptimal or surprise task conditions that can undermine success in real world settings. While we consider that function allocation may be appropriate for the more clearly structured simple problems, we argue that it is not a general solution that will remain robust in the face of complexity.

Here, in Figure 1, we offer an abstract illustration of how simple tasks require fundamentally different human-AI relationships than the complex tasks that AIs are likely to face in real-world application spaces. Our entry point is to discuss the task space of concern, which is that of human capability. We define this landscape as a map of normative human capability that varies as a function of the dynamic interaction along critical dimensions that are commonly used to differentiate human and AI strengths and weaknesses: time, information certainty, and complexity. Here, we depict the human capability map as time by information certainty cross sections from opposite extremes of the last dimension, complexity; we discuss each of these as follows.

The horizontal axis represents how much time is available to determine, formulate, and execute a response to influence the outcome of a given situation. Indeed, in an increasingly fast-paced world where computers are processing progressively more data more rapidly, time is believed to be an essential factor that differentiates appropriate task conditions for human versus automated inputs. Consider that, from a biological standpoint, human nerves and muscle tissues impose speed limits on initiating and executing any behavior; even ignoring perceptual and decision-making time requirements, basic human response times for simple movements take a minimum of 200 milliseconds [89]. When perceptual and decision-making times are included and real-world constraints are applied, response times grow to take seconds and beyond. For example, human drivers have been repeatedly observed requiring 1 – 1.5 seconds to engage the car's brake pedal in an emergency response, and consequently taking anywhere from 3 – 5 seconds to fully stop the vehicle [90], [91]. Given the very real and well-understood temporal limits of human response, time thresholds are frequently used to define situations in which AI can augment human capability.

The vertical axis represents the level of certainty in the information about the task and circumstance, including the amount of information available to formulate the best possible response. Commonly, a large amount of information uncertainty can spell deep trouble for an AI, as standard control systems tend to require well-defined, reliable inputs and analytically tractable objective functions. Intuitively then, humans become poised as final decision arbiters in complex

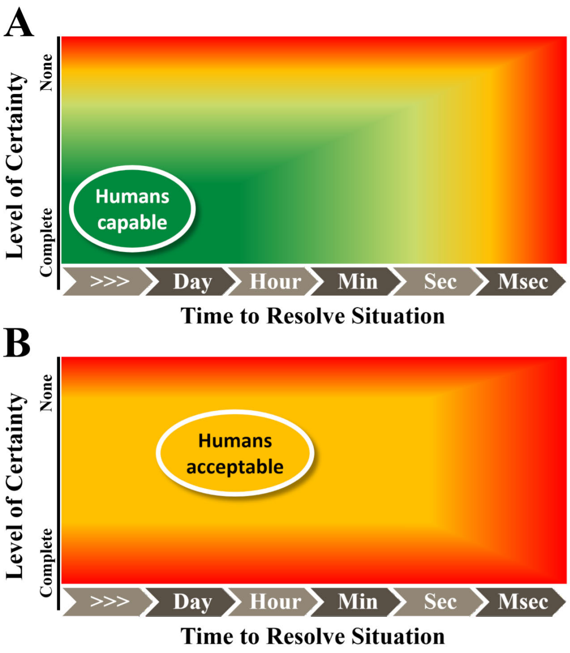


FIGURE 1. The landscape of human-AI partnership. This surface represents a map of capability as a function of information certainty and time required to resolve the situation. The two sub-panels represent cross-sections taken from opposite ends of the complexity dimension, which is a third axis that is not otherwise depicted here. Panel (A) shows human capability for solving problems within the simple domain, which we define as well-structured tasks that are bounded, require limited data, and may be solved with relatively common analytical tools. The simple domain includes clear-cut procedural tasks in which it is easy to grade performance, such as assembly line production. We consider these tasks as appropriate to consider from a function allocation perspective. Panel (B) shows human capability in the complex domain, which contains tasks that are more ambiguously structured in that they are either effectively or actually unbounded, involve large amounts of data, and cannot be solved analytically in polynomial time. Examples of complex problems include driving a car through city traffic during a storm, or formulating and executing a course of action to save lives and property during a house fire.

situations that are expected to challenge AI. However, significant uncertainty is similarly likely to force humans to adopt different methods for generating a response as well. Nevertheless, humans are often expected to be able to resolve ambiguities that may paralyze automated systems – such as in myriad designs for well-known human-robot interaction methods, including collaborative, supervisory, shared, traded, and similar control authority management schemes [92].

Finally, the third dimension, task complexity, is depicted here categorically (i.e., as panels) rather than continuously along an axis. Figure 1 represents this dimension as two separate panels drawn from extremes of the complexity dimension, with Panel A showing the human capability map for well-structured simple tasks, and Panel B showing the same for complex tasks. While the first two dimensions have received considerable attention in both human and AI research, and human-AI teaming specifically, it is this complexity dimension that we argue has been too often ignored in favor of more simplified concepts like function allocation. Nevertheless, the real-world is where most human-AI teams will need to function, and that context is, in our opinion, *defined* by complexity rather than simplicity and, moreover, is where suboptimal, ambiguously structured conditions are the norm rather than the exception.

In Figure 1, the color gradient represents human capability; in this case, that is the degree to which the human is capable of, and by extension should be responsible for, taking action on goals defined within the shared task space. As such, we consider this the nominal capability map for humans; though we note that the map is expected to vary quite dramatically within and across individuals as well as under dynamic performance constraints. We could envision similar landscapes that describe capability maps for a variety of different AI types, but for the moment focus on the human (theoretical capability maps are shown for several simple automations later in the paper). For any given task, such a map may be useful for ascribing so-called ''fiduciary responsibility'' [93], or what we would call privilege [13], for task performance.

Figure 1A depicts human capability to solve simple problems. The figure shows that, given sufficient time and information, humans are capable of solving simple problems (green portions of the space). With less time, or the human is not provided with enough certainty, performance is expected to suffer (red portions), and in the space between (yellow/orange), human capability is variable. We suggest that extending this logic to more complex problems risks mis-characterizing human capability with respect to these same dimensions. For complex problems, human capability does not necessarily improve given more time and a greater level of certainty (Figure 1B). Performance overall suffers with increasing complexity. However, presenting additional information does not always improve humans' ability to solve complex problems. Humans can become overwhelmed by large amounts of complex data (for example, see lay phenomena such as ''information overload'' [94]). In the following two sections, we provide evidence to substantiate the capability maps shown in Figure 1 above, in order to frame our discussion about role expectations in human-AI partnership.

B. SIMPLIFICATION HOLDS IN THE SIMPLE DOMAIN

Tasks in the simple domain can be characterized as well structured by clear temporal and spatial boundaries, limited data requirements, and are amenable to tractable analytical solutions. As depicted in Figure 1A, simple problems are easily addressed given sufficient information and time. Within this domain, we might increase difficulty by reducing the amount of the time permitted to solve the problem, or providing less information to reduce certainty. In such tasks, it is quite feasible to identify roughly linear relationships between time, certainty, and capability that can be used to guide human-AI partnership. We argue that Human–or–AI function allocation approaches only do well to characterize such simple human-AI partnerships.

Machines significantly outperform humans in solving simple problems. AI typically beats humans in terms of raw speed, particularly for simple operations; humans can solve simple multiplication problems in seconds [95] whereas AI can produce answers to these same problems in milliseconds. As computational power and AI approaches continue to advance, it is apparent that these technologies will be used to solve simple problems where speed is required. In terms of speed, the first oversimplification described above appears to hold up; AI may *replace* rather than integrate with humans for certain simple tasks, effectively operating as an advanced set of tools to perform basic tasks more quickly.

For simple problems, we also expect a linear relationship between information certainty and human capability. Decreasing certainty can create situations in which human cognitive biases degrade performance independent of the time scale. In the absence of certainty, humans frequently rely on heuristics as an adaptive way to reduce or restructure problems into forms that are more readily consumable. However, though often beneficial, these heuristics can also cause humans to make errors in judgment. For instance, it has been shown repeatedly that humans tend to misjudge the likelihood of low probability events [79], especially if those events are very consequential (e.g. different biases cause people to either over- or under-estimate the risk of chemical, biological, or nuclear terrorism [96]). People likewise fail to understand probabilities associated with independent events (as is the case in the ''hot hands'' phenomenon in which people falsely detect ''streaks'' of outcomes in random events [97]), and overestimate the probability of specific instances relative to general ones (i.e., the conjunction fallacy; [98]). One example of a simple but highly-uncertain problem is estimating the results of a sequence of fair coin tosses. Even in such a simple task, human decision making is subject to bias (i.e., the gambler's fallacy [99]), whereby people typically believe that if the coin shows tails more frequently than heads, it will eventually show more heads than tails in the near future. That is, human biases can powerfully override even the explicit knowledge that each coin toss event is independent and does not change the odds for successive flips. Moreover, as problems entail higher degrees of uncertainty, even simple predictive models that are far less capable than modern AI may be able to beat humans. To wit, in *Thinking Fast and Slow*, Kahneman [77] indicates that simple rules and algorithms often outperform human decision making for these types of problems because heuristics cannot be effectively deployed without also generating biased decisions.

Summarily, when simple challenges must be resolved quickly (e.g. milliseconds or less), or when they involve extremely high uncertainty, it is clear and reasonable to assign tasks to *either* a human or an AI, much like the HABA-MABA problem formulation that Fitts [72] suggested. Implicitly, however, this also suggests that simple domain problem solving *can always be improved by using a faster processor or actuator*, especially given that the types of problems that exist within this domain are generally analytical or, at the very least, can be expressed in terms of a known probability structure or control laws. Therefore, we concede that, in the simple domain, basic AI or, more likely AI-related technologies, can and often should be implemented as a replacement for human time and effort; to increase the speed and accuracy of problem solving, or to augment human decision making to mitigate biases. Comparatively, non-augmented human decision making will always be bounded by biological constraints in terms of time, which we argue is more relevant in this simple domain, and subject to bias. However, in the following section we argue that humans have evolved mechanisms that allow them to rapidly produce viable solutions even to complex problems, and therefore these oversimplified notions that appear to hold in the simple domain should not be expected to generalize well to the complex domain.

C. SIMPLIFICATION FAILS IN THE COMPLEX DOMAIN

The relationships between certainty and time, as observed in the complex domain, depart from the linear relationships that characterize human performance in the simple domain as depicted in Figure 1A. We generally characterize these complex problems as more ambiguously structured by having uncertain boundaries, if any, across time and space, requiring massive amounts of data in order to obtain complete certainty, and are computationally intractable for common analytic solutions; these problems may not have singularly optimal solutions, because solving for a particular criterion can (and often does) reduce the quality of that same solution when judged against other legitimate, yet competing goals (i.e., in a sufficiently complex, mixed-initiative system [100], all solutions reflect trade-offs). Human performance in the complex domain, as in the simple domain, is unlikely to yield successful, or reliable results when informational certainty, available time to respond, or both are low. Yet, unlike in the simple domain, attempting to obtain complete certainty about a complex problem may not improve performance. That is, obtaining complete certainty about a complex problem may require simultaneously processing so much high dimensional information as to be intractable, at which point working with it all becomes computationally infeasible.

One well-studied problem that demands reduction of informational complexity is the Traveling Salesman Problem (TSP). In this problem, the solver attempts to plan the shortest route through a set of nodes, representing cities, beginning and ending on the same node. The TSP is an NP-hard problem that is notoriously difficult for AI to solve through brute force. Humans produce near-optimal solutions to this problem in roughly linear time so long as the problem is presented visually, but when the problem is presented *as it is presented to computers* (i.e., as a distance matrix or matrix of coordinates from a rectangular *x-y* plane), human performance is substantially degraded [101]. In this case, humans achieve this high level of performance by leveraging the biologically- and psychologically-defined structure of the visual system to deploy a suitable cognitive strategy called reframing; that is, subdivide the problem into more manageable portions (i.e., local and global processing [102]) and then proceed working within them. More generally, human performance on such complex problems depends on the extent to which the problem is presented, or can be reformatted, to permit such processing. Contrary to the simple domain, the application of such cognitive heuristics – which we previously discussed as potentially manifesting as problematic biases – are here shown to also be highly adaptive and beneficial.

Complex problems tend to frustrate brute force solutions and require dimensional reduction so that solutions may be tractable. Many real-world problems, such as the TSP, are not mathematically reducible to polynomial time solutions [53]. An important characteristic of these problems is that increasing processing power alone will not produce transformational gains in a computer's ability to solve them. Barring potential approaches that might enable machines to access to human-like cognitive heuristics, this also then undermines human-AI partnering solutions that rely solely on function allocation, as they also collapse under the weight of combinatorics. For instance, consider the explosion in options that results within the TSP, where estimated solution time can increase dramatically, as in multiple orders of magnitude, relative to the more limited expansion of the search area (e.g. optimal route through 10 cities computed in milliseconds, through 15 cities required hours, and through 25 – *billions of years* [103]). Further, while increasing complexity reduces the "band" in which humans produce acceptable, but not necessarily optimal, solutions (see Figure 1B), they typically remain capable of generating multiple potential solutions relatively quickly without needing exact calculations. Humans accomplish this either by reframing, or reformatting, the problem in reduced dimension so that it is fit for their mental consumption, or they adapt by learning causal inference through repeated exposure and development of domain-specific expertise. In the next two paragraphs, we contrast examples of these problem solving approaches. Such strategies can disrupt the linear relationship between time and certainty and allow people to solve hard problems that challenge both humans and AI.

The aforementioned use case describes the Traveling Salesman Problem as one example of a task that is generally easy for humans to solve when presented in the visual modality. For other problems where certainty is much lower, human capability *can* be developed, learned, and optimized for success. Over the course of a lifetime, humans can develop the expertise necessary to make quick decisions in highly-complex yet uncertain situations [104]. However, expertise takes an extremely long time to develop and typically requires thousands or tens of thousands of hours of deliberate, repeated attempts to perform successfully [105] – a timeline that varies according to contributions of so-called natural abilities, individual proclivities, and other uniquely-experienced combinations of environmental support factors [106]. Expertise for very complex operational tasks, such as in military, hazmat, search-and-rescue, and others, generally must occur in high-risk domains as well, where the human cannot fail gracefully and risks considerable loss for a learning opportunity. The nature of the complex domain may not support life-long training for all human teammates, since the demand for those teammates in large numbers is high.

Understanding real-world complex problems can require an incomprehensibly vast amount of information. A 2014 report by RAND provided such an example, highlighting the U.S. Navy's big data challenges. The report concluded that the sheer number of sensors in the field, and amount of data collected, overwhelmed intelligence analysts and resulted in backlogs of information that might otherwise be actionable [107]. In this case, the problem was not a lack of certainty but rather a massive information overload that precluded timely human analysis. Similarly, AI systems are challenged by such data, which consists of signals intelligence across a wide array of informational modalities. Present day AI systems are designed to infer correlations, but do not have sufficient general intelligence to infer causality from events, or to combine and apply prior knowledge effectively to novel circumstances. Inferring causality from big data requires that the data be framed and reformatted such that the dimensionality is comprehensible, and humans are capable of doing this given sufficient time to conduct analysis.

An important practical and ethical challenge that is unique to human-AI interactions in the complex domain is the issue of safety, as well as the hard task of assigning responsibility for safety. Generally, we trust automation with our lives and livelihoods when the problems lie in the simple domain. For example, we do not typically second guess the outputs of our calculators during tax season and few regular air transit customers tend to regularly think twice about likely use of automated flight controls; the common element in both of these cases is the relative ease of separately defining human and AI responsibilities. However, humanity is still struggling to reach consensus on an effective risk calculus when it comes to integrating AI into the complex domain experiences dispersed throughout the daily lives of non-expert end users. As non-experts entrust AI-enabled agents with increasingly critical tasks, we must ask ourselves honestly about the level of proficiency that we expect: how safe is safe enough? Self-driving cars provide a good contemporary use case for answering this question. While individual risk calculus for autonomous vehicle safety may vary as a function of demographic factors like gender or nationality [108], estimates indicate that hundreds of millions, or even hundreds of billions, of miles of sample data would be required to demonstrate adequate safety margins [109]. Such estimates suggest that integration, assessment, and delegation of AI in complex problem spaces may need to be simultaneous and continuous. In practice, the company Tesla employed such a strategy – according to Bloomberg, testing their ''full self-driving'' capability first on volunteer employees in 2018 [110] and then on users [111] a year later. Of course, whether society widely consents to this approach remains to be seen.

Summarily, complex domain problems are ambiguously structured and offer challenges to both AI and humans. AI systems struggle to solve complex problems because they are either computationally intractable or require causal inference and reasoning skills that might be described in human-like terms, such as 'intuition'. Human performance, though, is bounded by the inherent capabilities of their individual biological systems – such as how time requirements may preclude certain human responses, or there may simply be too much data for a single human to process on demand. We have discussed how aspects of complex problems may remain solvable by humans or even require human intelligence to solve. However, relying on human intelligence is no silver bullet. Complex information may require advanced presentation and data exploration capabilities that enable re-formatting so the human can deploy heuristics or expertise that was learned over a lifetime of experience. Therefore, presently *neither* human or AI have an inherent or distinct advantage, we argue, for solving complex real-world problems. In the next section, we further suggest that exclusive Human–or–AI approaches will not provide the robust solutions that we seek for problems in the complex domain. Rather, we expect to find solutions in an intelligent sociotechnical ecosystem that exploits a stable set of solutions, or approaches that specify particular interaction modalities, that coexist within the landscape of human-AI partnership.

IV. INTELLIGENT SOCIOTECHNICAL ECOSYSTEMS

A main take-away of the discussion to this point is that humans and AIs are highly interdependent, especially when it comes to joint task work in the complex domain. This, we believe, is true regardless of whether any of the involved agents has learned (or been developed) to recognize their interdependence with the other(s). Moreover, the defining characteristics of this reciprocal interdependence can vary widely enough that a broad ecosystem of human-AI interaction dynamics will be required to support a similar breadth of joint system performance capabilities.

A. PRINCIPLED BEGINNINGS

In previous work, we developed a principle-driven framework for guiding decisions about the structure of formal control laws to instantiate a variety of interaction dynamics (The Privileged Sensing Framework, or PSF [13]). The intention of the PSF was to provide a relatively simple, yet principled and generalizable approach for integrating inputs from a heterogeneous human-AI system across a broad variety of team configurations and performance requirements. The primary mechanism for blending control authority was weighting inputs from various system agents on the basis of a computed quantity denoted as ‘‘privilege’’; here specified as a dynamic variable based on quantitative estimates of confidence (e.g. uncertainty) and consequence (e.g. risk-reward ratio) associated with the observed states of each agent, as well with as other aspects of the task environment. Articulated in a small set of principles, the PSF was highly adaptable, and it was implemented in a number of formalized control structures, such as a standard weighted sum model and a novel Dynamic Belief Fusion method. Through intelligent human-AI partnerships, PSF-based control systems successfully improved performance across a varied set of tasks that included remote asset path planning, target tracking and engagement, control authority designation during semi-automated driving, and rapid image triage [13], [112], [113].

While we found the PSF useful as a simple, general, and scaleable approach for fusing inputs into a control decision in a limited number of cases, we also consider it fairly mute as to when, or in what circumstances, one would choose any particular allocation of privilege across a team. That is, in order to determine the optimal system configuration and design appropriate control laws, one must first understand the structure and/or allocation of privilege across the human-AI team for that task. In what parameter regimes is the human most likely to perform superior to a robotic counterpart or vice-versa? Under what conditions will a particular operator be likely to under-perform or interrupt in potentially catastrophic ways? What is the cost of switching or even sharing task authority across the team? These and other questions lie at the core of determining (or even understanding) the inherent privilege structure for a task, given the capabilities of the various agents in the human-AI team. We also believe that the concept of privilege is compatible with the landscape of human-AI partnership as discussed here, and further, may be informed to a great extent by the specific capability maps for each of the agents in the system. We spend the remainder of this section discussing a simplified example, and conclude this with a section discussion implications for more complex, real-world challenges and demands.

Figure 2 provides an explicit, though theoretical, example of how a currently well-known human-machine partnership may be better understood as an ecosystem of intelligent agents. For this example, we consider the human driver of a standard commuter car that has three minimally-intelligent (though not actual AI), automated driver support systems; these include a blind spot monitoring system (BSMS), an automatic emergency braking system (AEBS), and an intelligent navigation assistant (NAV); the corresponding capability maps for each of these agents and a human are shown in panel A on the left. Each of these theoretical maps has the same structure as defined in Figure 1, with the vertical axis representing certainty and the horizontal axis is available time to respond. The capability map for the human (top) is the same as that in Figure 1B, showing moderate human capability exists in a broad band throughout the complex domain, but also with significant performance reduction or increased variability as expected with too little time available to respond as well as either too much information or too little certainty.

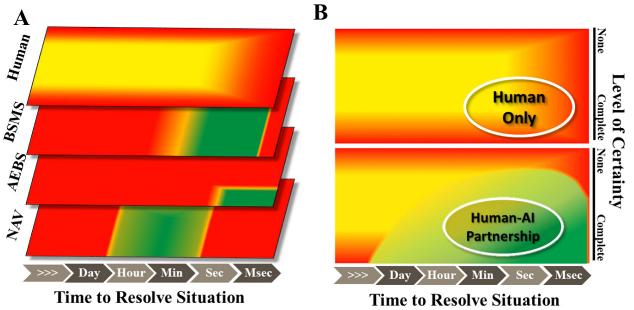


FIGURE 2. A simplified example of an ecosystem that could support enhanced driving of a roadway vehicle. The left panel (A) shows the capability maps for the human driver, a blind spot monitoring system (BSMS), an automatic emergency braking system (AEBS), and an intelligent navigation system (NAV) and the right panel (B) contrasts the theoretical human-only capability map with one representing a composite of the human-AI partnership in this ecosystem. As with Figure 1, all charts show capability as a function of information certainty (vertical axes) and time available to respond (horizontal axes); the red, yellow, green color map corresponds with low, moderate, and high capability, respectively.

For the remaining agents in this example, we have envisioned support automations that each have a high performance capability that is nevertheless restricted to narrow portions of the landscape. The capability maps shown in Figure 2 reflect that they, like most technologies, are made for specific purposes and will only work properly when used in a particular way and are generally useless otherwise. The BSMS, for instance, is envisioned here to operate well even under relatively high uncertainty, which is what we expect given the system's purpose. As a warning system, a blind spot monitor would be wisely designed be over-permissive in triggering alerts; i.e., it is not nearly as harmful to look thrice before a lane change than to not look at all. The AEBS, on the other hand, should only activate when conditions unambiguously indicate an imminent collision without sufficient time for a human to act effectively. The NAV system, as we envision, would have a broader operating range in terms of both certainty and available time, but it only has the highest performance (darkest green) in a very narrow strip around the time scale of minutes. This is because traffic and weather conditions are rather dynamic; so, the accuracy and precision of estimates diminish with shorter or longer time intervals.

Panel B on the right is intended to show the impact of the ecosystem on improving capability in a very global way, at least in the portion of the landscape that is relevant to our driving example here. Causally, the difference between the human-only map on the top (and the individual maps in panel A) and the joint human-AI map on the bottom is that each agent supports and elevates the capabilities of the ecosystem as a whole. Importantly, the capability map for the human-AI ecosystem does not result from a summation or convolution of the four individual agent capability maps. Rather, the whole is greater than the sum of the parts because of emergent capabilities that arise *only* because of unique interactions that exist between and within various subspaces of the ecosystem wherein subsets of agents drive specific interactions and behaviors. Consider, for instance, how the human and the NAV system mutually improve one another and, ideally, result in selection and completion of the most optimal path through a given area, whereas each may have made suboptimal choices along the way if left to their own calculations or judgements, given their particular capabilities and limitations.

B. REAL-WORLD IMPLICATIONS

Given the expected nonlinear expansion of complexity across the landscape as real-world ecosystems will include larger and more heterogeneous groups, we posit that the most effective system designs will be dominated by those that enable understanding, characterizing, and leveraging the dynamically interactive nature of human-AI partnerships, as in our ecosystems conceptualization above. We are not the first to arrive at these kinds of conclusions, as a variety of others have offered concepts and frameworks that are compatible with a human-AI ecosystem's approach as discussed herein. Examples include more holistic, systems-level design approaches like Rasmussen's Ecological approach [114] and the Joint-Cognitive Systems approach taken by Woods and colleagues [17], mechanisms for human-robot collaboration (e.g. [115]–[118]), and, most importantly, approaches that eschew hierarchical or centralized control in favor of polycentric architectures (e.g. [13], [119]–[121]). With continued success in developing AI and AI-related technologies that can engage more fully with the complexities of the real world, we must adopt something broader than the *user – tool* mentality; here, we advocate for taking the vision of intelligent sociotechnical ecosystems as inspiration for innovating and manifesting true human-AI partnership.

Ecosystems in the real-world, of course, will be far from Utopian. Much of the trade-space in human-AI partnership already receives regular attention, but is usually discussed in more limited contexts; the latest model self-driving car versus the model that just had a fatal accident, the most recently trending deep fake video versus the enhanced facial recognition security on your phone – indeed, the examples are plentiful. We devote a minimum of space here to discuss several aspects of the trade-space involved and yet we ultimately argue that the collective will always have a broader and more robust capability set than the sum of the individual agents working independently [55].

Control frameworks designed for collectives of human-AI partnership need not be heavy-handed in assigning authority within a fixed regime. Rather such partnerships may provide a literal menu of controllable and adaptable human interventions for complex and uncertain challenges. Of course, many complex circumstances benefit from the scrutiny of a human mind and its associated inductive reasoning capabilities; and yet, this inherently means that such a circumstance also may need to be handled at a slower pace than if processed entirely digitally. Nevertheless, herein lies the opportunity for a flexible human-AI partnership. Because, while even a highly qualified AI may not be able to fully resolve the complexity, it very likely could rapidly generate a limited set of potential solutions along with projected performance estimates based on a model that was pre-specified and vetted through human processes like test and evaluation. Such centaur teams have been shown to outperform both humans and AIs in games, such as chess [122], and have shown promise in real-world domains such as medical decision making [123], mission planning [124], and cybersecurity [125].

Just as with the complexity, a trade space also exists for human and AI responses with respect to information certainty. Despite the progress noted in the discussion of oversimplifications above, current generation AI continues to face difficulty in identifying general categories that are critical for dealing with novelty [6]. For instance, humans do not usually need specific training (or re-training) in order to be able to intuit things like danger and risk in novel situations. AIs, on the other hand, can mislabel novel situations as something they have observed before and therefore respond in potentially maladaptive ways [126]. Human responses in highly uncertain environments tend to be stereotyped and made within the context of survival, while AI responses tend to appear random and inscrutable. An implication of this is that the AI behavior may not match human expectations because of differences in the AI's underlying reasoning process, and humans' mental models of it. As a result, AIs may fail in ways that human teammates do not expect [127], and produce solutions that, though optimal, differ from the preferences of human teammates [128]. These disparities risk fracturing the human-AI partnership – whether in the form of misuse, disuse, or abuse [129]. Recent efforts to train AIs using human-assisted machine learning (e.g., [130]–[132] are susceptible to similar issues, as it is possible that a robot trained by one human may not produce the exact behavior that is either predictable or preferred by other human teammates. Similarly, there is no guarantee that behaviors learned in this way would generalize to novel contexts in ways that would match human expectations. Human mental models are particularly mercurial when it comes to emerging technologies, and can be influenced by irrelevant factors such as superficial morphological features [133], [134] and prior experience that may not be relevant (for review, see [135]).

Finally, human-AI partnership in the complex domain requires a substantial investment in terms of creating the types of systems that support interactions for appropriately calibrating and maintaining trust to preserve compatibility, mutual acceptance, and ultimately, teamwork for effective performance [136]. The trade-space is indeed complex and must be carefully considered. At the same time, as we show in Figure 3, if the trade space is navigated adequately and we learn how to both encourage and exploit emergent capabilities across our intelligent sociotechnical ecosystems, then we believe that we can manifest a landscape marked by broad capability to function effectively in complex contexts. In the final section of this paper, we attempt to ground this discussion a little more firmly in real-world examples with a set of practical, implementable use-cases, presented as a bit of a tour through one manifestation of an intelligent sociotechnical ecosystem for human-AI partnership.

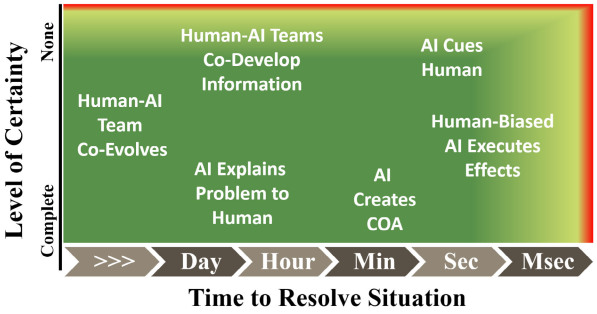

Level of Certainty
None
Complete
Human-AI Teams Co-Develop Information
AI Cues Human
Human-AI Team Co-Evolves
Human-Biased AI Executes Effects
AI Explains Problem to Human
AI Creates COA
>>>
Day
Hour
Min
Sec
Msec
Time to Resolve Situation

FIGURE 3. **Here we illustrate that a set of human-AI partnerships that have been purposefully designed for the constraints within their region of the landscape, when brought together as a collective, may provide the best potential for maintaining effectiveness across nearly the entire range of simple and complex tasks; the only exception being at the very margins of physical and informational tractability.**

V. USE CASES IN THE JOINT HUMAN-AI ECOSYSTEM

In this final section, we illustrate a set of use cases as examples of situations wherein adopting different interaction strategies may effectively expand the envelope of capability within intelligent sociotechnical ecosystems. The overall take-away is that human-AI partnerships, if fitting the task constraints defined by certainty, time, and complexity, may provide for broad capability enhancement as compared with traditional Human–or–AI function allocation methods that steadfastly maintain, if not tacitly assume, independent and isolated roles for each agent type. Here, six general types of interaction strategies are discussed vis-à-vis use cases (A through F) that are meant to illustrate how each strategy may operate under various task contexts. We begin discussion on the far right of the time axis (Figure 3), where available response windows are the smallest, and then progressively explore how human-AI partnering may vary as we travel across the landscape. While a variety of technologies are presented, their function is to illustrate the value of the interaction strategy; any perceived endorsement of specific devices or technologies is not intended beyond the evidentiary value they hold for demonstrating real-world application.

A. HUMAN-BIASED AI EXECUTES EFFECTS

This partnering strategy is likely to be the best and only way to enable human influence to be accounted for in circumstances where time is too short to allow for querying or otherwise awaiting human feedback and input. In such cases, which often involve significant safety and/or security objectives (e.g. split-second auto-braking to avoid a sudden traffic accident), it is conceivable that the human can pre-specify a ‘‘bias’’ for how the AI should behave before the situation arises. For example, modern adaptive cruise control technologies allow drivers to set a preferred following distance as well as the desired set-point speed, presumably to engender greater trust through having the vehicle behave more consistently with the person’s own preferred driving style; a similar concept could be implemented for ‘‘braking aggressiveness.’’ Another example of this type of partnership today includes the option for human users to configure their network firewall protection levels based on how they want to balance productivity against security risk. In general, humans can understand the broader context, but are too slow to be effective in the near instantaneous decision-making loops required for effectiveness in the cyber domain. Because humans fundamentally understand the problem space, however, they can effectively make advance judgements to set bias parameters enabling the AI confidence in selecting responses while avoiding the inherent reductions to efficiency if needing to seek even occasional human approval at run time.

B. AI CUES HUMAN

Indeed, there are increasingly complex circumstances where human attention may improve the overall team response, but that attention is divided and the time available to formulate that response is too abbreviated to allow much room for judgement and consideration of alternatives. In circumstances where the levels of certainty and time are low, then an AI that is capable of detecting patterns in noisy data may have an important role in providing alerts and suggestions. Yet, in such time- and certainty-limited situations, the human may also have an important role in supporting the AI. In this portion of the landscape, the AI may be fully capable of observing an emerging pattern within a multi-dimensional data set and successively concluding that circumstances are concerning, but may still be unable to converge on an objectively preferred course of action (COA). In such cases, the AI-based system may be imbued with decision rules that indicate a need to very quickly draw the human into the problem space, and the human may consequently be in the best position to make an authoritative decision about the COA. Given that a human may not always be available, a timeout function may also be included that would execute some default action if there is no human response within a configurable, pre-defined response window (e.g. as used in a process control simulation studied by Moray and colleagues [137]). An example of this type of human-AI partnership today is a modern intelligent security system that can identify suspicious circumstances and then, based on particular decision criteria, selectively cue the human to look at a video and ultimately make a decision as to whether to contact authorities or ignore the alert (e.g., commercial home door automations, like Ring™or Google Nest®, that can sense a potential security issue and then send a video message from a doorbell camera to the homeowner). In general, humans are able to understand the broader context and, given time, are able to draw conclusions based on sparse data more effectively than the AI that may not be fully likely to track relevant context cues.

C. AI CREATES COA

In the case where time available to respond remains at the level of a few seconds to a few minutes (as with the previous use case B), but there is much greater certainty in the task-relevant information, the AI may take a more active and assertive role in developing courses of action for the human to evaluate. We envision this interaction strategy would be most appropriate in situations where time is very limited, but the AI is capable of identifying the circumstance with enough certainty that it can develop satisfactory COAs. Nevertheless, it may still be desirable or necessary for the human to make the final selection. Here, the AI may best be used to quickly draw the person into a common problem space through a salient alert signal, and then recommend several alternatives that reflect the risks, trade-offs, and projected outcomes associated with each COA, presented with confidence bounds to facilitate human trust and confidence. Examples of this type of human-AI partnership today include car navigation systems that can present several alternative routes and provide information about their characteristics. Other examples include mixed-initiative decision making where the AI proposes COAs and the human selects among them (e.g., collaborative human-automation scheduling of multiple unmanned vehicles [124]). Critically, in complex, big data situations, the AI would be able to process much more information than the human in the limited time available; here, humans adopt a more supervisory role. This type of human-AI partnership will only be accepted over the long term, however, if a strong bi-directional trust can indeed be manifest between the human and the AI [138].

D. HUMAN-AI TEAMS CO-DEVELOP INFORMATION

As the time constraints become less immediately limiting, we start to see the role of the human shift towards engaging the AI "in the loop," even in mechanistic ways. This use case is characterized as a situation of response windows ranging from hours to days and beyond, where human or AI teammates conclude there may be an issue, but neither the human nor the AI can select an optimal, or even satisfactory, solution on their own because of a relatively low amount of information certainty. More to the point, humans and AIs are likely to achieve different degrees of certainty for similar aspects of a decision – meaning that their strengths and weaknesses may be complementary within limits. In such cases, the value of joint, collaborative processing by both human and AI will enable development of greater certainty for more robust performance. Relatively current examples where human-AI teams have outperformed both human experts and specialized AIs include: *FOLDIT*, a human-directed computing approach to protein folding where humans propose solutions that are successively evaluated by an AI (c.f. [139]) and *Centaur chess*, where the human chess players make decisions in collaboration with AI that can process, store, and recall tens of millions of chess matches (c.f. *The Average is Over* [140]). In the finance world where high performing human-AI partnerships have been developed for financial decision making (e.g., [141]), one company has gone so far as to elect an AI to its board of directors, allowing it one of six votes on investment strategy (c.f. Yuval Noah Harari, *Homo Deus: A Brief History of Tomorrow* [142], p. 437). Again, joint problem solving has typically found utility in cases where the AI and the human each are well-suited for addressing different and limited aspects of the overall task and, by working together, can improve performance across multiple criteria. For instance, such a partnering strategy may be useful for cases where it is critical to find a balance involving performance trade-offs across multiple objectives (e.g. response speed vs accuracy; [13]). There are certain problems that AIs are not expected to solve in a satisfactory manner on their own. These problems are referred to as AI-hard or AI-complete, an equivalency drawn to NP-hard and NP-complete problems in computational complexity theory. An AI-hard problem has been defined formally as a problem that at least a subset of humans can solve given unlimited time, is composed of a set of instances and a probability distribution over that set, and for which verifiably correct answers are available [143]. For complex problems that are both AI-hard and difficult for humans, human-AI teams may be able to solve them through co-development of information.

E. AI EXPLAINS PROBLEM TO HUMAN

When the time available to respond remains on the order of hours to days and informational certainty is high, meaning that solutions are identifiable, known, and/or may be chosen with confidence, we would expect the human-AI relationship to appear like that of competent and skilled individuals (teammates) working together on a common problem. In this portion of our landscape, an AI may have to develop explanations that allow the human to understand the problem sufficiently to adjudicate and provide feedback on recommendations as to which conclusions and actions are critical. We see big data problems as having a home in this category, as there would simply be too much data for the human to process and comprehend in any amount of time. Yet, AI naturally excels with big data, and especially when given enough processing time and power, can find new, interesting, and relevant patterns in the data that the human might otherwise overlook or never even encounter. For these types of problems, humans will remain unlikely to understand the deep specifics of various multidimensional feature spaces that form the substance of the AI's model. However, human judgment may still be required or beneficial in forming a decision. Therefore, it is here that explainable AI becomes ever more important. In order for the humans to believe that the solution is valid, the AI needs to exhibit transparency by providing a degree of explanatory insight into how it arrived at its conclusions, or at least what information contributed to them, in what way, and to what extent. These situations may be useful for improving trust in the AI and may be particularly well-suited for overall strategy and team development, for instance during training and after-action reviews. With better established trust, the human may be more readily willing to partner with the AI when the time is limited and the human is not be able to fully understand the problem and solution space (see Use Cases A and B, in particular).

F. HUMAN-AI TEAM CO-EVOLVES

Over the course of extremely long timescales, the individual and collaborative behaviors of agents in human-AI teams will naturally tend to mutually adapt as they meet the changing context of the world in which they work. This is because both humans and AIs are fundamentally learning agents that are capable of evolving over their entire lifespans. This natural co-adaptation or co-evolution will be rather advantageous to enable individuals and collectives to remain competitive. For the AI, adaptation will be needed ensure that it does not become outmoded or overcome with changes in context and complexity. Over what are essentially developmental time scales, humans may provide machine-interpretable explanations about their own performance as well as with respect to that of the AI, as in recent work on human-guided reinforcement learning [42]. The AI, likewise, may also take task execution data along with the human feedback and extract meaning in a format that enables the AI to evolve itself and tune its responses to those that result in better team function, cohesion, and communication. We believe that developments within this portion of the landscape can lead to a massive increase, perhaps up to 100 fold, in the ability of complex, intelligent sociotechnical ecosystems to co-evolve across their lifecycles – from initial formation, through situation-based training, and then through the operational life cycle as individual agents and teams perform and consequently mature together [23]. On the longest timescales, this also represents the means of evolving the team such that it can dynamically select interaction strategies from across the entire ecosystem. We expect this flexibility to autonomously change interaction methods will provide an array of human-AI partnership options to expand the envelope of performance potential, and further enhance the reliability and robustness of the entire ecosystem.

VI. CONCLUSION

AI and AI-related technologies are rapidly evolving, broadening in scope of application, and supporting societal advances within both technology-driven and developing societies around the globe. Yet, we contend that the world has only begun to see the dramatic ways in which lives are likely to change. We have argued here that an important limiting factor on the depth to which AI and AI-related technologies are accepted and integrated as part of society is that, in the main, many have not yet shifted their mindset about true nature of change that AI can bring – at least not beyond the polarity of, on one hand, very simplified concepts like AI taking working-class jobs or, on the other hand, highly unrealistic scenarios borne from science fiction. There remain many oversimplified ways of considering where and how AI will intersect with and influence the very nature of human activity in both constructive and destructive ways. In the present paper, we have argued that the best way to go beyond these limitations is by ceasing to consider AI as simply a tool and, instead, come to new understandings of what happens when AI-based technologies are treated as potential partners with whom collaborative mechanisms may change depending upon the task and context – just like they do for exclusively-human teams. We believe that this notion of AI-as-teammate (or partner), taken into consideration while accounting for factors of information *certainty*, available *time* to respond, and task *complexity*, is a core aspect of the current revolution that is unfolding before our eyes. Old, basic notions such as Human–or–AI task assignment and function allocation will no longer provide tenable methods to support the complex, adaptive, intelligent sociotechnical ecosystems that are emerging across all sectors of human society. Rather, such simplified concepts must give way to new paradigms of human-AI partnership; the risk of not doing so, we contend, may involve creating the exact future that many are hoping to avoid.

ACKNOWLEDGMENT

(Jason S. Metcalfe and Brandon S. Perelman contributed equally to this work.) The authors of this article would like to acknowledge, with gratitude, discussions with Dr. Philip Perconti regarding human-AI role allocation as inspiration for the development of the concepts within this article.

REFERENCES

[1] A. Kott, "Toward universal laws of technology evolution: Modeling multi-century advances in mobile direct-fire systems," *J. Defense Model. Simul.*, vol. 17, no. 4, pp. 373–388, Oct. 2020, doi: 10.1177/1548512919875523.

[2] R. Kurzweil, *The Singularity Is Near: When Humans Transcend Biology*. Menlo Park, CA, USA: Google, Sep. 2005.

[3] J. N. Pelton, *Preparing for Next Cyber Revolution: How Our World Will Be Radically Transformed Again*. Cham, Switzerland: Springer, 2019. [Online]. Available: http://link.springer.com/10.1007/978-3-030-02137-5

[4] A. Curioni, "Artificial intelligence: Why we must get it right," *Informatik-Spektrum*, vol. 41, no. 1, pp. 7–14, Feb. 2018, doi: 10.1007/s00287-018-1087-0.

[5] A. Holzinger, P. Kieseberg, E. Weippl, and A. M. Tjoa, "Current advances, trends and challenges of machine learning and knowledge extraction: From machine learning to explainable AI," in *Machine Learning and Knowledge Extraction* (Lecture Notes in Computer Science), A. Holzinger, P. Kieseberg, A. M. Tjoa, and E. Weippl, Eds. Cham, Switzerland: Springer, 2018, pp. 1–8.

[6] G. Marcus, "Deep learning: A critical appraisal," 2018, *arXiv:1801.00631*. [Online]. Available: http://arxiv.org/abs/1801.00631

[7] J. Kantor, "Man and machine in science," *J. Philosophy*, vol. 32, no. 25, pp. 673–684, 1935.

[8] H. Birmingham and F. Taylor, "A design philosophy for man-machine control systems," *Proc. IRE*, vol. 42, no. 12, pp. 1748–1758, Dec. 1954.

[9] R. C. Goertz and W. M. Thompson, "Electronically controlled manipulator," *Nucleonics (US) Ceased publication*, vol. 12, p. 15, Oct. 1954.

[10] C. E. Billings. (Aug. 1999). *Human-Centered Aircraft Automation: A Concept and Guidelines*. [Online]. Available: https://ntrs.nasa.gov/search.jsp?R=19910022821

[11] M. M. Cummings, "Man versus machine or man + machine?" *IEEE Intell. Syst.*, vol. 29, no. 5, pp. 62–69, Sep. 2014.

[12] S. W. A. Dekker and D. D. Woods, "MABA-MABA or abracadabra? Progress on human-automation co-ordination," *Cognition, Technol. Work*, vol. 4, no. 4, pp. 240–244, Nov. 2002, doi: 10.1007/s101110200022.

[13] A. R. Marathe, J. S. Metcalfe, B. J. Lance, J. R. Lukos, D. Jangraw, K.-T. Lai, J. Touryan, E. Stump, B. M. Sadler, W. Nothwang, and K. McDowell, "The privileged sensing framework: A principled approach to improved human-autonomy integration," *Theor. Ergonom. Sci.*, vol. 19, no. 3, pp. 283–320, May 2018, doi: 10.1080/1463922X.2017.1297865.

[14] W. D. Nothwang, M. J. McCourt, R. M. Robinson, S. A. Burden, and J. W. Curtis, "The human should be part of the control loop?" in *Proc. Resilience Week*, Aug. 2016, pp. 214–220.

[15] R. Parasuraman and C. D. Wickens, "Humans: Still vital after all these years of automation," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 50, no. 3, pp. 511–520, Jun. 2008, doi: 10.1518/001872008X312198.

[16] T. B. Sheridan, "Function allocation: Algorithm, alchemy or apostasy?" *Int. J. Hum.-Comput. Stud.*, vol. 52, no. 2, pp. 203–216, Feb. 2000. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1071581999902859

[17] D. D. Woods, "Cognitive technologies: The design of joint human-machine cognitive systems," *AI Mag.*, vol. 6, no. 4, p. 86, 1985. [Online]. Available: https://www.aaai.org/ojs/index.php/aimagazine/article/view/511

[18] D. D. Woods and M. Branlat, "Hollnagel's test: Being 'in control' of highly interdependent multi-layered networked systems," *Cognition, Technol. Work*, vol. 12, no. 2, pp. 95–101, Jun. 2010, doi: 10.1007/s10111-010-0144-5.

[19] Y. H. Yin, A. Y. C. Nee, S. K. Ong, J. Y. Zhu, P. H. Gu, and L. J. Chen, "Automating design with intelligent human machine integration," *CIRP Ann.*, vol. 64, no. 2, pp. 655–677, Jan. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S000785061500147X

[20] J. Aagaard-Hansen, "The challenges of cross-disciplinary research," *Social Epistemol.*, vol. 21, no. 4, pp. 425–438, 2007.

[21] M. Monteiro and E. Keating, "Managing misunderstandings: The role of language in interdisciplinary scientific collaboration," *Sci. Commun.*, vol. 31, no. 1, pp. 6–28, Sep. 2009.

[22] J. M. Bradshaw, P. Feltovich, M. Johnson, M. Breedy, L. Bunch, T. Eskridge, H. Jung, J. Lott, A. Uszok, and J. van Diggelen, "From tools to teammates: Joint activity in human-agent-robot teams," in *Proc. Int. Conf. Hum. Centered Design*. Berlin, Germany: Springer, 2009, pp. 935–944.

[23] A. H. DeCostanza, A. R. Marathe, A. Bohannon, A. W. Evans, E. T. Palazzolo, J. S. Metcalfe, and K. McDowell, "Enhancing human-agent teaming with individualized, adaptive technologies: A discussion of critical scientific questions," U.S. DEVCOM Army Res. Lab., Aberdeen Proving Ground, MD, USA, Tech. Rep. ARL-TR-8359, May 2018. [Online]. Available: https://apps.dtic.mil/sti/citations/AD1051552

[24] C. J. Thoman, "Sir humphry davy and frankenstein," *J. Chem. Educ.*, vol. 75, no. 4, p. 495, Apr. 1998.

[25] I. Asimov, *I, Robot*. Greenwich, CT, USA: Fawcett Publications, 1950.

[26] J. M. Keynes, *Economic Possibilities for Our Grandchildren*. London, U.K.: Springer, 2010.

[27] E. McGaughey, "Will robots automate your job away? Full employment, basic income, and economic democracy," Centre for Bus. Res., Univ. Cambridge, U.K., Tech. Rep. 496, 2018.

[28] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, and O. Evans, "Viewpoint: When will AI exceed human performance? Evidence from AI experts," *J. Artif. Intell. Res.*, vol. 62, pp. 729–754, Jul. 2018.

[29] A. Smith and J. Anderson, "AI, robotics, and the future of jobs," *Pew Res. Center*, vol. 6, p. 51, Oct. 2014.

[30] T. Walsh, "Expert and non-expert opinion about technological unemployment," *Int. J. Autom. Comput.*, vol. 15, no. 5, pp. 637–642, Oct. 2018.

[31] Y. Ivanenko and V. S. Gurfinkel, "Human postural control," *Frontiers Neurosci.*, vol. 12, p. 171, Mar. 2018. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2018.00171/full

[32] A. Fuchs and J. S. Kelso, "Coordination dynamics and synergetics: From finger movements to brain patterns and ballet dancing," in *Complexity and Synergetics*. Cham, Switzerland: Springer, 2018, pp. 301–316.

[33] C. Johnson, "Self-driving cars will have to decide who should live and who should die. Here's who humans would kill," The Washington Post, Washington, DC, USA, Oct. 2018. [Online]. Available: https://www.washingtonpost.com/science/2018/10/24/self-driving-cars-will-have-decide-who-should-live-who-should-die-heres-who-humans-would-kill/

[34] T. K. Adams, "Future warfare and the decline of human decision making," *US Army War College Quart., Parameters*, vol. 41, no. 4, p. 1, 2011.

[35] Y. Li, C. Burns, and R. Hu, "Understanding automated financial trading using work domain analysis," *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 59, no. 1, pp. 165–169, Sep. 2015.

[36] E. Board. (2020). *Knight Blowup Shows How High-Speed Traders Outrace Rules*. [Online]. Available: https://www.bloomberg.com/opinion/articles/2012-08-06/knight-blowup-shows-how-high-speed-traders-outrace-rules

[37] A. Kharpal. (2017). *Elon Musk: Humans Must Merge With Machines or Become Irrelevant in AI Age*. Accessed: Feb. 13, 2017. [Online]. Available: http://www.cnbc.com/2017/02/13/elon-musk-humans-merge-machines-cyborg-artificial-intelligence-robots

[38] E. Musk and Neuralink, "An integrated brain-machine interface platform with thousands of channels," *J. Med. Internet Res.*, vol. 21, no. 10, Oct. 2019, Art. no. e16194.

[39] A. Impacts. (2020). *Brain Performance in TEPS*. [Online]. Available: https://aiimpacts.org/brain-performance-in-teps/

[40] J. Hsu, "Estimate: Human brain 30 times faster than best supercomputers," *IEEE Spectr.*, Aug. 2015. [Online]. Available: https://spectrum.ieee.org/tech-talk/computing/networks/estimate-human-brain-30-times-faster- than-best-supercomputers

[41] B. Kim and B. Pardo, "A Human-in-the-Loop system for sound event detection and annotation," *ACM Trans. Interact. Intell. Syst.*, vol. 8, no. 2, pp. 1–23, Jul. 2018.

[42] N. R. Waytowich, V. G. Goecks, and V. J. Lawhern, "Cycle-of-Learning for autonomous systems from human interaction," 2018, *arXiv:1808.09572*. [Online]. Available: http://arxiv.org/abs/1808.09572

[43] D. J. Power and R. Sharda, "Model-driven decision support systems: Concepts and research directions," *Decis. Support Syst.*, vol. 43, no. 3, pp. 1044–1061, Apr. 2007.

[44] T. Karmakharm and P. Richmond, "Large scale pedestrian multi-simulation for a decision support tool," in *Proc. TPCG*, 2012, pp. 41–44.

[45] J. Dzieza. *How Hard Will the Robots Make US Work*. Accessed: Feb. 9, 2021. [Online]. Available: https://www.theverge.com/2020/2/27/21155254/automation-robots-unemployment-jobs-vs-human-google-amazon

[46] K. Leetaru. *As AI Comes for Management Perhaps we Should Look Forward to Machines Taking Our Jobs*. Accessed: Feb. 9, 2021. [Online]. Available: https://www.forbes.com/sites/kalevleetaru/2019/06/24/as-ai-comes-for-management-perhaps-we-should-look-forward-to-machines-taking-our-jobs/?sh=55b69117624d

[47] P. Sindawi. *The Boss Machine is Here—Ai is All Set to Eliminate Middle Management in 8 Years*. Accessed: Feb. 9, 2021. [Online]. Available: https://www.businessinsider.in/careers/news/the-boss-machine-is-here-ai-is-all-set-to-eliminate-middle-managers-in-8-years/articleshow/73474729.cms

[48] C. Pazzanese. *Ethical Concerns Mount as AI Takes Bigger Decision-Making Role in More Industries*. Accessed: Feb. 9, 2021. [Online]. Available: https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/

[49] E. Sherman. *AI Could Take Over Management Jobsif Artificial Intelligence can Make Better Decisions, Why Employ a Human Manager*. Accessed: Feb. 9, 2021. [Online]. Available: https://www.inc.com/erik-sherman/your-next-manager-could-be-a-computer.html

[50] D. Tobenkin. *HR Needs to Stay Ahead of Automation*. Accessed: Feb. 9, 2021. [Online]. Available: https://www.shrm.org/hr-today/news/hr-magazine/spring2019/pages/hr-needs-to-stay-ahead-of-automation.aspx

[51] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Math.*, vol. 1, no. 1, pp. 269–271, Dec. 1959.

[52] P. Hart, N. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE Trans. Syst. Sci. Cybern.*, vol. 4, no. 2, pp. 100–107, Oct. 1968.

[53] R. M. Karp, "Reducibility among combinatorial problems," in *Complexity of Computer Computations*. Boston, MA, USA: Springer, 1972, pp. 85–103.

[54] N. C. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "The computational limits of deep learning," 2020, *arXiv:2007.05558*. [Online]. Available: http://arxiv.org/abs/2007.05558

[55] T. W. Malone, *Superminds: Surprising Power People Computing Thinking Together*. New York, NY, USA: Little, Brown and Company, 2018,

[56] K.-F. Lee. *Artificial Intelligence is Powerful-and Misunderstood. Here's How we Can Protect Workers*. Accessed: Feb. 16, 2021. [Online]. Available: https://time.com/5501056/artificial-intelligence-protect-workers/

[57] C. B. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to computerisation?" *Technol. Forecasting Social Change*, vol. 114, pp. 254–280, Jan. 2017.

[58] B. Goertzel, "Artificial general intelligence: Concept, state of the art, and future prospects," *J. Artif. Gen. Intell.*, vol. 5, no. 1, pp. 1–48, Dec. 2014.

[59] R. J. Sternberg, *Intelligence*. Hoboken, NJ, USA: Wiley, 2013.

[60] E. R. Kandel, J. H. Schwartz, T. M. Jessell, D. of Biochemistry, M. B. T. Jessell, S. Siegelbaum, and A. Hudspeth, *Principles of neural science*, vol. 4. New York, NY, USA: McGraw-Hill, 2000.

[61] D. Chalmers, "The hard problem of consciousness," in *Proc. Blackwell Companion Consciousness*, 2007, pp. 225–235.

[62] R. A. Khalil, E. Jones, M. I. Babar, T. Jan, M. H. Zafar, and T. Alhussain, "Speech emotion recognition using deep learning techniques: A review," *IEEE Access*, vol. 7, pp. 117327–117345, 2019.

[63] B. Ko, "A brief review of facial emotion recognition based on visual information," *Sensors*, vol. 18, no. 2, p. 401, Jan. 2018.

[64] M.-H. Huang and R. T. Rust, "Artificial intelligence in service," *J. Service Res.*, vol. 21, no. 2, pp. 155–172, 2018.

[65] M. Mazzone and A. Elgammal, "Art, creativity, and the potential of artificial intelligence," *Arts*, vol. 8, no. 1, p. 26, Feb. 2019.

[66] J. B. Voytek and B. Voytek, "Automated cognome construction and semi-automated hypothesis generation," *J. Neurosci. Methods*, vol. 208, no. 1, pp. 92–100, Jun. 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027012001513

[67] M. Cherti, B. Kegl, and A. Kazakci, "Out-of-class novelty generation: An experimental foundation," in *Proc. IEEE 29th Int. Conf. Tools with Artif. Intell. (ICTAI)*, Nov. 2017, pp. 1312–1319.

[68] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robot. Auto. Syst.*, vol. 57, no. 5, pp. 469–483, May 2009.

[69] M. Conti. *The Incredible Inventions of Intuitive AI*. [Online]. Available: https://www.ted.com/talks/maurice_conti_the_incredible_inventions_of_intuitive_ai

[70] N. Kato, H. Osone, D. Sato, N. Muramatsu, and Y. Ochiai, "DeepWear: A case study of collaborative design between human and artificial intelligence," in *Proc. 12th Int. Conf. Tangible, Embedded, Embodied Interact.*, Mar. 2018, pp. 529–536.

[71] L. V. Ahn, "Games with a purpose," *Computer*, vol. 39, no. 6, pp. 92–94, Jun. 2006.

[72] P. M. Fitts, Ed., *Human Engineering for an Effective Airnavigation and Traffic-Control System* (Human Engineering for an Effective Air-Navigation and Traffic-Control System). Oxford, U.K.: National Research Council, 1951.

[73] J. M. Bradshaw, V. Dignum, C. Jonker, and M. Sierhuis, "Human-agent-robot teamwork," *IEEE Intell. Syst.*, vol. 27, no. 2, pp. 8–13, Mar. 2012.

[74] J. C. F. de Winter and D. Dodou, "Why the fitts list has persisted throughout the history of function allocation," *Cognition, Technol. Work*, vol. 16, no. 1, pp. 1–11, Feb. 2014, doi: 10.1007/s10111-011-0188-1.

[75] A. De Santis, B. Siciliano, A. De Luca, and A. Bicchi, "An atlas of physical human–robot interaction," *Mech. Mach. Theory*, vol. 43, no. 3, pp. 253–270, Mar. 2008.

[76] M. A. Goodrich and A. C. Schultz, *Human-Robot Interaction: A Survery*. New York, NY, USA: Now, 2008.

[77] D. Kahneman, *Thinking, Fast and Slow*. New York, NY, USA: Macmillan, 2011.

[78] G. Gigerenzer and W. Gaissmaier, "Heuristic decision making," *Annu. Rev. Psychol.*, vol. 62, pp. 451–482, 2011.

[79] A. Tversky and D. Kahneman, "Judgement under uncertainty: Heuristics and biases," *Science*, vol. 185, pp. 1124–1134, Oct. 1974.

[80] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises," *Rev. Gen. Psychol.*, vol. 2, no. 2, pp. 175–220, Jun. 1998.

[81] V. Hoorens, "Self-enhancement and superiority biases in social comparison," *Eur. Rev. Social Psychol.*, vol. 4, no. 1, pp. 113–139, Jan. 1993.

[82] M. D. Alicke and O. Govorun, "The better-than-average effect," *Self Social Judgment*, vol. 1, pp. 85–106, 2005.

[83] R. Chesney and D. K. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," *Calif. L. Rev.*, vol. 107, p. 1753, 2019.

[84] D. Heaven, "Why deep-learning AIs are so easy to fool," *Nature*, vol. 574, no. 7777, pp. 163–166, Oct. 2019.

[85] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno, and D. Song, "Robust physical-world attacks on deep learning visual classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1625–1634. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2018/html/Eykholt_Robust_Physical-World_Attacks_CVPR_2018_paper

[86] S. T. Mueller, "Cognitive anthropomorphism of AI: How humans and computers classify images," *Ergonom. Des., Quart. Hum. Factors Appl.*, vol. 28, no. 3, pp. 12–19, Jul. 2020, doi: 10.1177/1064804620920870.

[87] M. A. Alcorn, Q. Li, Z. Gong, C. Wang, L. Mai, W.-S. Ku, and A. Nguyen, "Strike (with) a pose: Neural networks are easily fooled by strange poses of familiar objects," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4845–4854.

[88] A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2015, pp. 427–436.

[89] D. L. Woods, J. M. Wyma, E. W. Yund, T. J. Herron, and B. Reed, "Factors influencing the latency of simple reaction time," *Frontiers Human Neurosci.*, vol. 9, p. 131, Mar. 2015.

[90] G. M. Fitch, M. Blanco, J. F. Morgan, and A. E. Wharton, "Driver braking performance to surprise and expected events," *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 54, no. 24, pp. 2075–2080, Sep. 2010.

[91] S. Young Sohn and R. Stepleman, "Meta-analysis on total braking time," *Ergonomics*, vol. 41, no. 8, pp. 1129–1140, Aug. 1998.

[92] F. Flemisch, D. A. Abbink, M. Itoh, M.-P. Pacaux-Lemoine, and G. Weßel, "Joining the blunt and the pointy end of the spear: Towards a common framework of joint action, human–machine cooperation, cooperative guidance and control, shared, traded and supervisory control," *Cognition, Technol. Work*, vol. 21, no. 4, pp. 555–568, Nov. 2019.

[93] B. M. Muir, "Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems," *Ergonomics*, vol. 37, no. 11, pp. 1905–1922, Nov. 1994, doi: 10.1080/00140139408964957.

[94] B. M. Gross, *The Managing of Organizations*. Chicago, IL, USA: Free Press of Glencoe, 1964.

[95] I. Imbo, A. Vandierendonck, and Y. Rosseel, "The influence of problem features and individual differences on strategic performance in simple arithmetic," *Memory Cognition*, vol. 35, no. 3, pp. 454–463, Apr. 2007, doi: 10.3758/BF03193285.

[96] G. D. Koblentz, "Predicting peril or the peril of prediction? Assessing the risk of CBRN terrorism," *Terrorism Political Violence*, vol. 23, no. 4, pp. 501–520, Sep. 2011.

[97] T. Gilovich, R. Vallone, and A. Tversky, "The hot hand in basketball: On the misperception of random sequences," *Cognit. Psychol.*, vol. 17, no. 3, pp. 295–314, Jul. 1985. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0010028585900106

[98] A. Tversky and D. Kahneman, "Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment," *Psychol. Rev.*, vol. 90, no. 4, pp. 293–315, 1983.

[99] A. Tversky and D. Kahneman, "Belief in the law of small numbers," *Psychol. Bull.*, vol. 76, no. 2, p. 105, 1971.

[100] K. Drnec, G. Gremillion, D. Donavanik, J. D. Canady, C. Atwater, E. Carter, B. A. Haynes, A. R. Marathe, and J. S. Metcalfe, "The role of psychophysiological measures as implicit communication within mixed-initiative teams," in *Proc. Int. Conf. Virtual, Augmented Mixed Reality*. Cham, Switzerland: Springer, 2018, pp. 299–313.

[101] N. I. Polivanova, "Functional and structural aspects of the visual components of intuition in problem solving. [Functional and structural aspects of the visual components of intuition in problem solving.]," *Voprosy Psychologii*, vol. 4, pp. 41–51, Oct. 1974.

[102] X. Kong and C. D. Schunn, "Global vs. Local information processing in visual/spatial problem solving: The case of traveling salesman problem," *Cognit. Syst. Res.*, vol. 8, no. 3, pp. 192–207, Sep. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389041707000216

[103] O. Abdoun, J. Abouchabaka, and C. Tajani, "Analyzing the performance of mutation operators to solve the travelling salesman problem," 2012, *arXiv:1203.3099*. [Online]. Available: http://arxiv.org/abs/1203.3099

[104] G. A. Klein, "A recognition-primed decision (RPD) model of rapid decision making," in *Decision Making in Action: Models and Methods*. Westport, CT, USA: Ablex, 1993, pp. 138–147.

[105] K. Ericsson, R. Krampe, and C. Tesch-Romer, "The role of deliberate practice in the acquisition of expert performance," *Psychol. Rev.*, vol. 100, no. 3, pp. 363–406, 1993.

[106] D. Z. Hambrick and R. R. Hoffman, "Expertise: A second look," *IEEE Intell. Syst.*, vol. 31, no. 4, pp. 50–55, Jul. 2016.

[107] I. Porche, B. Wilson, E.-E. Johnson, E. Saltzman, and S. Tierney, *Data Flood: Helping the Navy Address the Rising Tide of Sensor Information*. Santa Monica, CA, USA: Rand Corporation, Apr. 2014.

[108] J. Moody, N. Bailey, and J. Zhao, "Public perceptions of autonomous vehicle safety: An international comparison," *Saf. Sci.*, vol. 121, pp. 634–650, Jan. 2020.

[109] N. Kalra and S. M. Paddock, "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?" *Transp. Res. A, Policy Pract.*, vol. 94, pp. 182–193, Dec. 2016.

[110] D. Hull and J. Eidelson. *Tesla Enlists Employees to be 'Full Self-Driving' Beta Testers*. Accessed: Mar. 7, 2021. [Online]. Available: https://www.bloomberg.com/news/articles/2018-09-28/tesla-enlists-employees-to-be-full-self-driving-beta-testers

[111] Z. Mider. *Tesla's Autopilot Could Save The Lives of Millions, But it Will Kill Some People First*. Accessed: Mar. 7, 2021. [Online]. Available: https://www.bloomberg.com/news/features/2019-10-09/tesla-s-autopilot-could-save-the-lives-of-millions-but-it-will-kill-some-people-first

[112] A. R. Marathe, B. T. Files, J. D. Canady, K. A. Drnec, H. Lee, H. Kwon, A. Mathis, W. D. Nothwang, G. Warnell, and E. Stump, "Heterogeneous systems for information-variable environments (hive)," U.S. DEVCOM Army Res. Lab., Aberdeen Proving Ground, MD, USA, Tech. Rep. ARL-TR-8027, 2017.

[113] J. Metcalfe, "Building a framework to manage trust in automation," *Proc. SPIE Micro Nanotechnol. Sensors, Syst., Appl.*, vol. 10194, Oct. 2017, Art. no. 101941U.

[114] J. Rasmussen, "Ecological Interface Design for Reliable Human-Machine Systems," *The Int. J. Aviation Psychol.*, vol. 9, no. 3, pp. 203–223, Jul. 1999, doi: 10.1207/s15327108ijap0903_2.

[115] D. J. Bruemmer, D. A. Few, R. L. Boring, J. L. Marble, M. C. Walton, and C. W. Nielsen, "Shared understanding for collaborative control," *IEEE Trans. Syst., Man, Cybern. A, Syst. Humans*, vol. 35, no. 4, pp. 494–504, Jul. 2005.

[116] P. O. Dusadeerungsikul and S. Y. Nof, "A collaborative control protocol for agricultural robot routing with online adaptation," *Comput. Ind. Eng.*, vol. 135, pp. 456–466, Sep. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360835219303675

[117] T. Fong, C. Thorpe, and C. Baur, "Multi-robot remote driving with collaborative control," *IEEE Trans. Ind. Electron.*, vol. 50, no. 4, pp. 699–704, Aug. 2003.

[118] S. Y. Nof, "Collaborative control theory for E-work, E-production, and e-Service," *Annu. Rev. Control*, vol. 31, no. 2, pp. 281–292, Jan. 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1367578807000430

[119] G. Grote, "Uncertainty management at the core of system design," *Annu. Rev. Control*, vol. 28, no. 2, pp. 267–274, Jan. 2004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1367578804000331

[120] G. Grote, J. Weyer, and N. A. Stanton, "Beyond human-centred automation – concepts for human–machine interaction in multi-layered networks," *Ergonomics*, vol. 57, no. 3, pp. 289–294, Mar. 2014, doi: 10.1080/00140139.2014.890748.

[121] E. Hollnagel, D. D. Woods, and N. Leveson, *Resilience Engineering: Concepts and Precepts*. Farnham, U.K.: Ashgate, 2006.

[122] P. D. Scharre, "The opportunity and challenge of autonomous systems," in *Autonomous Systems: Issues for Defence Policymakers*. Norfolk, VA, USA: NATO Supreme Allied Command Transformation, 2006, pp. 3–26.

[123] I. M. Goldstein, J. Lawrence, and A. S. Miner, "Human-machine collaboration in cancer and beyond: The centaur care model," *JAMA Oncol.*, vol. 3, no. 10, pp. 1303–1304, Oct. 2017. [Online]. Available: https://jamanetwork.com/journals/jamaoncology/fullarticle/2599994

[124] A. S. Clare, "Modeling real-time human-automation collaborative scheduling of unmanned vehicles," Ph.D. dissertation, Dept. Aeronaut. Astronaut., Massachusetts Inst. Technol., Cambridge, MA, USA, 2013.

[125] K. Veeramachaneni, I. Arnaldo, V. Korrapati, C. Bassias, and K. Li, "AI^2: Training a big data machine to defend," in *Proc. Int. Conf. Big Data Secur. Cloud BigData Secur.*, Apr. 2016, pp. 49–54.

[126] B. M. Lake and M. Baroni, "Generalization without systematicity: On the compositional skills of sequence-to-sequence recurrent networks," 2017, *arXiv:1711.00350*. [Online]. Available: http://arxiv.org/abs/1711.00350

[127] R. R. Hoffman, S. T. Mueller, G. Klein, and J. Litman, "Metrics for explainable AI: Challenges and prospects," 2018, *arXiv:1812.04608*. [Online]. Available: http://arxiv.org/abs/1812.04608

[128] B. S. Perelman, A. W. Evans III, and K. E. Schaefer, "Where do you think You're going?: Characterizing spatial mental models from planned routes," *ACM Trans. Hum.-Robot Interact.*, vol. 9, no. 4, pp. 1–55, Oct. 2020.

[129] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 39, no. 2, pp. 230–253, Jun. 1997.

[130] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proc. 21st Int. Conf. Mach. Learn.* Banff, AB, Canada: Association for Computing Machinery, Jul. 2004, p. 1, doi: 10.1145/1015330.1015430.

[131] D. Ramachandran and E. Amir, "Bayesian inverse reinforcement learning," in *Proc. 20th Int. Joint onf. Artif. Intell.* Hyderabad, India: Morgan Kaufmann, Jan. 2007, pp. 2586–2591.

[132] B. D. Ziebart, A. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning," in *Proc. 23rd Nat. Conf. Artif. Intell.*, Chicago, IL, USA: AAAI Press, Jul. 2008, pp. 1433–1438.

[133] S. Kiesler and J. Goetz, "Mental models of robotic assistants," in *Extended Abstracts on Human Factors in Computing Systems*. Minneapolis, MN, USA: Association for Computing Machinery, Apr. 2002, pp. 576–577, doi: 10.1145/506443.506491.

[134] E. Salas, D. E. Sims, and C. S. Burke, "Building an ethical culture" *Small Group Res.*, vol. 36, no. 5, pp. 555–599, Oct. 2005, doi: 10.1177/1046496405277134.

[135] E. Phillips, S. Ososky, J. Grove, and F. Jentsch, "From tools to teammates: Toward the development of appropriate mental models for intelligent robots," *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 55, no. 1, pp. 1491–1495, Sep. 2011.

[136] K. E. Schaefer, B. S. Perelman, G. M. Gremillion, A. R. Marathe, and J. S. Metcalfe, *Trust in Human Robot Interaction*. Oxford, U.K.: Academic, 2021.

[137] N. Moray, T. Inagaki, and M. Itoh, "Adaptive automation, trust, and self-confidence in fault management of time-critical tasks," *J. Experim. Psychol., Appl.*, vol. 6, no. 1, pp. 44–58, 2000.

[138] M. J. Ashleigh and N. A. Stanton, "Trust: Key elements in human supervisory control domains," *Cognition, Technol. Work*, vol. 3, no. 2, pp. 92–100, Apr. 2001.

[139] S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popoviá, and F. Players, "Predicting protein structures with a multiplayer online game," *Nature*, vol. 466, no. 7307, pp. 756–760, Aug. 2010.

[140] T. Cowen, *Average Is Over: Powering America Beyond the Age of the Great Stagnation*. Menlo Park, CA, USA: Google, 2013.

[141] Y. Nagar and T. W. Malone, "Improving predictions with hybrid markets," in *Proc. AAAI Fall Symp. Series*, Oct. 2012, pp. 1–5. [Online]. Available: https://www.aaai.org/ocs/index.php/FSS/FSS12/paper/view/5653

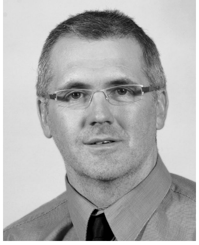
[142] Y. N. Harari, *Homo Deus: A Brief History of Tomorrow* (Vintage Popular Science). London, U.K.: Harvill Secker, 2016.

[143] L. von Ahn, M. Blum, N. J. Hopper, and J. Langford, "CAPTCHA: Using Hard AI Problems for Security," in *Advances in Cryptology* (Lecture Notes in Computer Science), E. Biham, Ed. Berlin, Germany: Springer, 2003, pp. 294–311.

JASON S. METCALFE received the B.S. and M.S. degrees from the University of Illinois at Urbana–Champaign and the Ph.D. degree from the University of Maryland, all in kinesiology. For more than 25 years, he has leveraged methods from developmental psychology, cognitive neuroscience, biomechanics, and engineering to parlay his understanding of human sensory-motor control into expertise on human performance in complex environments. He is currently a Research Kinesiologist with the U.S. DEVCOM Army Research Laboratory, where he also co-leads the Center for Agent-Soldier Teaming. He has published more than 45 works and his research has been the subject of more than 65 presentations; and notable awards, include the University of Maryland Doctoral Fellow, in 2003, the University of Maryland Distinguished Instructor, in 2007, and the ARL Human Research and Engineering Directorate Award for Science, in 2017. His current research interests include human behavioral prediction and trust-based decision-making in the context of human-intelligent agent interaction, with a particular emphasis on real-world interactions with intelligent vehicles.

BRANDON S. PERELMAN was born in Oregon City, OR, in 1985. He received the B.S. degree in biology and psychology from the Sarah Lawrence College, Bronxville, NY, USA, in 2007, the M.S. degree in experimental psychology from Saint Joseph's University, Philadelphia, PA, USA, in 2012, and the Ph.D. degree in applied cognitive science and human factors from Michigan Technological University, Houghton, MI, USA, in 2015. Prior to attending graduate school, he has served on the Israel Defense Forces' Nahal Brigade. From 2012 to 2016, he has worked as the Junior Cognitive Scientist with Applied Research Associates Inc., on research products for ARI, AFRL, IARPA, and FBI. In 2014, he completed the Summer Fellowship at AFRL, Wright-Patterson AFB. In 2015, he began a Postdoctoral Fellowship at DEVCOM ARL, Human Research and Engineering Directorate, where he has been working as the Research Psychologist, since 2019. His awards and honors include the AFRL's Daniel Repperger Fellowship, in 2014, the Michigan Technological University's Outstanding Student Teacher Award, in 2014, the Technical Publication Award for his work on IARPA's SIRIUS Program, in 2013, and ARL Awards for leadership and impactful communication, in 2019.

DAVID L. BOOTHE was born in Washington, DC, USA, in 1967. He received the B.A. degree in philosophy from the University of Maryland at College Park, College Park, in 1989, and the Ph.D. degree in computational neuroscience from the Neuroscience and Cognitive Sciences Program, University of Maryland, in 2007. He is currently a Researcher with the U.S. DEVCOM Army Research Laboratory. He specializes in neuroscience, simulation, and complex systems.



KALEB MCDOWELL (Senior Member, IEEE) was born in Frederick, MD, USA, in July 1970. He received the B.S. degree in operations research and industrial engineering from Cornell University, Ithaca, NY, USA, in 1992, and the M.S. degree in kinesiology and the Ph.D. degree in neuroscience and cognitive science from the University of Maryland at College Park, College Park, MD, USA, in 2000 and 2003, respectively. He is currently the Chief Scientist of the U.S. Army Research Laboratory's Human Research and Engineering Directorate. Since joining DEVCOM ARL, he has developed a strong record of publication and impact within government, industry, and academic research and development communities; and he has led several major research and development programs focused on neuroscience/neuroengineering, indirect vision systems, vehicle mobility, and human-agent teaming; and receiving Army Research and Development Achievement Awards, in 2007 and 2009, ARL Awards for Leadership and Engineering, in 2011 and 2013, and numerous Funding Awards.