June 2021

Design Considerations for Future Human-AI Ecosystems

Jason S. METCALFE [a,1] and Jurriaan van DIGGELEN [b]
[a] *Humans in Complex Systems, DEVCOM Army Research Laboratory, United States*
[b] *Human-Machine Teaming, TNO, The Netherlands*

Abstract. This paper is intended to open a discussion with those who engage in research and technology development for advanced artificial intelligence (AI) of the future; particularly those who seek to manifest a net-positive impact on society. Most critically, the design and deployment of advanced human-AI interactions must unfold as a continuous and highly interactive process, mirroring the nature of the future intelligent sociotechnical ecosystems that are envisioned. Importantly, we expect traditional roles such as designer and end-user become less distinct, and believe that blurring these lines and fostering increasing complex interaction dynamics may present opportunities to overcome persistent challenges in the domain.

Keywords. human-AI partnership, intelligent sociotechnical ecosystems, meaningful human control, human-centered design

1. Introduction

Our reality is shifting before our eyes, whether or not we can apprehend the full extent of change that is coming. Throughout history, much of our species development has either provoked or been precipitated by change in technology [1]. Things that once lived in the domain of science fiction are now just what we call science. Importantly, history teaches that as these broad shifts occur, we have a choice. We can passively accept, and thus become subject to the will of the larger ecosystem, or we can actively embrace reality and work to influence the how the change is realized. Our choice is the latter, and we contend that this choice brings particular moral and ethical imperatives to establish *meaningful human control* (MHC; [2]) of the artificial intelligence (AI) and AI-related technologies that are anticipated to pervade our future sociotechnical landscape. In this paper, we discuss human-centered design of partnerships with AI and related technologies, with an intention to support the long-term growth of intelligent sociotechnical (human-AI) ecosystems that produce a net-positive impact on society [3].

By specifying "meaningful" in MHC, we acknowledge that *responsibility* comes with authority. Through the establishment of MHC, then, we envision a sociotechnical reality where humans and AIs[2] carry this responsibility as they cooperate, compete, co-evolve, and ultimately support mutual growth, learning, and development as true part-

[1]Corresponding Author: Jason S. Metcalfe, US DEVCOM Army Research Laboratory, Aberdeen Proving Ground, MD, USA; E-mail: jason.s.metcalfe2.civ@army.mil

[2]For simplicity, we will shorten "AI and AI-related technologies" to "AI" for the rest of this paper

ners. Importantly, we note that MHC does not mean creating an exclusively human-dominated future – indeed myriad interaction paradigms exist to support human-AI collaboration to allow us to reap the benefits, such as increased productivity, safety, and quality [4, 5]. Critically, we argue that control mechanisms must evolve well-beyond function allocation-oriented approaches [6–8] owing to their tendency to draw sharp lines designating human-or-AI authority based on individual capability in the space of action. Instead, we believe that our vision of the future calls for more nuanced interleaving of individual authority with collective responsibility in the space consequences [3].

2. A New, Old Way of Addressing Pervasive Challenges

In the space of human-AI interaction, there are many options for drawing and enforcing lines between human and AI control [4, 8]. Nevertheless, there is a perplexing paucity of AIs that are actually well-integrated with human society at large; this, despite the mass proliferation of demonstrable "smart technologies" (e.g. phones, home automation, vehicles, etc). The list of potential causes for this gap is long, and yet a clear set of robust solutions remains elusive. Here, we offer that part of this circumstance may be attributable to *how* the challenges are being considered. In particular, we suggest that letting go of classic systems engineering thinking for a moment may provide some valuable insights. Design thinking has been part of user-centered design of interactive human-computer systems at least since Norman and Draper published their seminal volume in the 1980's [9], but it has not tended to be as strongly influential for implementation and evaluation of those same systems. While some have criticized design thinking as too esoteric to manifest macro-scale societal changes [10], we contend that it offers useful ways to encourage valuable re-thinking of our research thrusts. Two features of design thinking are especially useful here: adopting a *solution-focus* over a problem-focus and *reshaping* the design process to suggest new routes for addressing difficult challenges.

2.1. A Solution-Focus on Human-AI Partnership

In most instances within systems engineering, the problem tends to precede the solution as a main point of focus; particularly during early ideation aimed at producing a specific end-result. For example, until very recently, AI-related technologies have been mainly developed to address specific, independent needs (e.g. in transportation, social media, space exploration, stock trading, health care) and, in that, without deep consideration of possible impacts on one another. Naturally, this makes sense, because the standards in science and engineering encourage (if not demand) a narrow focus on a specific problem. Such practices are especially necessary and effective for generating reliable products within the typically tight, competition-sensitive timelines that are driven by modern science and technology ambitions. We contend that the strength of the problem-focus is, however, also its weakness. This is because the ultimate target of the knowledge and technical products is commonly an application within a larger sociotechnical ecosystem, and unanticipated cross-domain interactions can easily become the source of significant dysfunction. Social media, for instance, can and does affect the dynamics of the stock market [11] and has been witnessed globally to have influenced population-scale decision making. Here, we suggest that the time is ripe for application of solution-focused

design thinking; and the focus of this paper on ecosystem-scale concerns constitutes a demonstration of this kind of thinking.

2.2. Reshaping the Design Process to Match the Dynamic Product

Typically, identifying a problem or even just noting sub-optimal performance, can trigger a process that then unfolds in a series of discrete, or perhaps iterative, phases - beginning with design and ending with evaluation (see Figure 1).

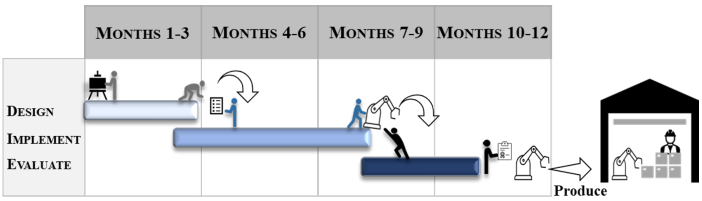

Figure 1. A depiction of the general technology development process as a Gantt chart that is augmented in cartoon-like fashion to represent the relationships between the individuals who support design, implementation, evaluation, and the end-user. Colors of the human silhouettes correspond with their role in the process.

The situation is critically different for AI and related technologies. With the increasingly intelligent and adaptive systems of the future, we and others contend that traditional approaches are ill-suited because of the capability of AI, and its human ”users”, to continuously learn and co-adapt. Because of this continuous mutual adaptation, necessary processes for the development of trustworthy systems cannot rely on traditional methods for design, implementation, and evaluation, but must instead be re-shaped into something that allows for continuous and flexible engagement of all stakeholders through the whole system lifecycle [12]. Importantly, in such a reshaped paradigm, the lines between ”designer”, ”evaluator”, and ”user” start to blur, where end-users (and eventually AIs) become both evaluators and designers as well.

Figure 2 leverages the classic ”Decision Ladder” from the work of Rasumussen and Goodstein [13] as the basis for a reshaped, dynamic decision cycle. This reshaped, dynamic decision cycle places evaluation centrally, rather than at the end of the process (blue, dashed rectangle). More importantly, as depicted here, evaluation is a critical interface between design (left side) and implementation (right side), as well as being a conduit of communication for all human and AI stakeholders. Similar cartoons as in Figure 1 are placed here in locations that represent how such interactions may be facilitated, such as the end user standing on the **Feedback** box at the bottom, which provides information to the designer, or at the level of the selected **Target** on the right depicting an interaction with an implementer who carries **Predict(ed) Consequences** from a cooperative (and AI augmented) analysis between designer and evaluator, as shown at the apex of the ladder. Moreover, this figure also represents a few different kinds of interactions between the human stakeholders and an AI, as with the robot in the evaluation loop at the top. In one case, the AI may autonomously shortcut the decision ladder based on relevant **Feedback** and **Observation** data, or the AI may take a shortcut to intervene and/or advise the designer-user team as they select a new **Target** based on that same data. Moreover, in a third shortcut, we see that a team of human designer and implementer may guide the AI towards appropriate **Task** for execution at the **Planning** phase (alternatively, one might interpret the AI as pushing the humans to the answer instead).

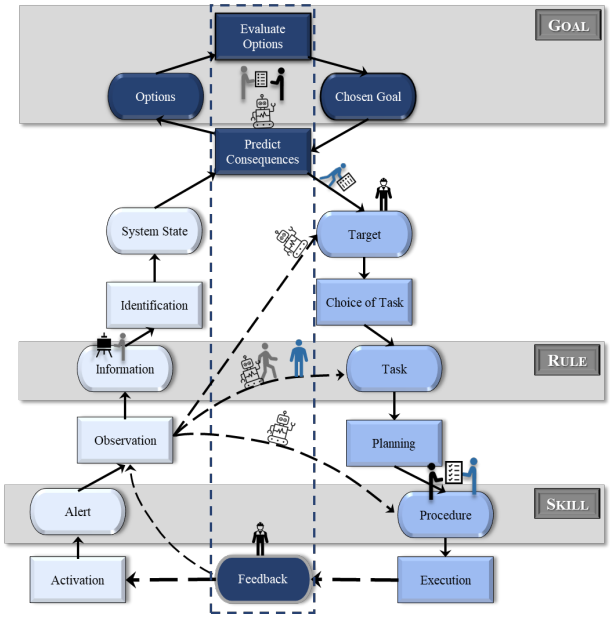

Figure 2. One representation of a reshaped design process which leverages Rasumssen and Goodstein's Decion Ladder. Here the ladder has been modified to close the loop with feedback from the execution to the next activation. Further, cartoons have been added to indicate some ways in which the lines become blurred between users and developers (see discussion in text). Color codes correspond with those used in Figure 1.

3. Design Considerations for Complex Coordination and Delegation

By fostering the development of a completely interactive ecosystem, we invoke a need for a language that facilitates dynamic cooperation. Within social and cognitive sciences, this issue has been examined through the study of team performance; and within these studies, both individual and team dynamics have been recognized as uniquely important [14, 15]. *Taskwork* has been used to describe specific processes that each individual team member must engage to perform their job, while *teamwork* describes the multi-agent interactions needed for communication and coordination, as well as for developing team trust and cohesion [16]. Commonly, human-AI interactions are designed at one, but not both of these levels; though both are essential for effective team performance and establishing MHC. Here, we offer that the reshaped decision cycle incorporates aspects of both, and this positions the task as the common language of cooperation.

3.1. The Task as a Language: Goals, Rules, and Skills

Regardless of application, it is often a task that serves as impetus for developing a human-AI partnership. Therefore, an intelligent sociotechnical ecosystem will almost always be organized based on tasks, whether explicitly or implicitly. Tasks are goal-directed activities that must be conducted under some set of *rules*, and task-based *goals* may be achieved by application of skilled behaviors. As depicted in Figure 2, *skills* represent the most elemental level of organization in the decision cycle. Here, skills constitute taskwork, and they represent a level of capability that is typically defined in terms of

parameters that specify an effective action. Of course, having an ability to execute elemental skills is not enough to assure the kind of effective and adaptive performance that would be considered a hallmark of an intelligent ecosystem [17]; those skills should be performed at particular moments and in ways that conform with the needs of the task. Understanding when and how to perform skills is therefore in the domain of rules, which can be seen in Figure 2 as defined at an informational level within the decision cycle. Given their position in the decision cycle, rules also provide for the interface between taskwork (skills) and teamwork (communication and cooperation). Finally, goals represent the highest level of organization and planning, and the selection of goals is often a deliberative process involving consideration of options, simulating or otherwise predicting outcomes of actions, and then selecting the goals that will be most likely to manifest desired consequences. While goals may often be chosen by a singular commander or supervisor, our depiction of the decision cycle is meant to reflect that they may also be identified, evaluated, and chosen through collective decision making.

3.2. A Solution Focus: Designing Delegation for Meaningful Human Control

MHC is more than a framework to guide human-AI team configuration and authority delegation for optimal performance in the space of actions. MHC is meant to enable the responsible selection of a course of action to produce a specific outcome, as well as to foster and propagate individual and shared accountability, which may then drive future improvements through learning. All actions have consequences, and designers cannot feasibly predict all possible outcomes or complications. Sometimes, chosen actions will have unanticipated effects. Better, in our minds, to consider how to develop and propagate the shared responsibility for all consequences regardless of whether they were intended. Some forms of delegating authority may be better suited for this final purpose than others; skill-based delegation really just assigns a particular task to the most suitable agent that is available, while rule-based delegation may be enabled through the use of pre-defined plays that specify particular patterns of both teamwork and taskwork, as well as the responsibilities and accountabilities that go with each. Both kinds of delegation allow passing of authority to different (human and/or AI) team members based on their capacity to perform the tasks as needed, and should, we argue, incorporate an expectation that each agent will participate in and support the teamwork needed for developing and maintaining the shared responsibility for the actions of the collective and their consequences. This will become even more important as continued advances lead to goal-based delegation, where humans and AIs alike can receive a request or command and then faithfully ensure it is achieved without requiring a human supervisor to outline all of the specific rules and skills that should be used in the process. Ultimately, as the intelligent ecosystems of the future become realities of the present, we will aim to define success less in terms of gain and loss and more in terms of our ability to produce desired consequences as well as to maintain responsibility and accountability while doing it.

4. Conclusion

As we began this paper, we noted that our overall objective is to better understand how to use human-AI partnership to manifest net-positive impact in future sociotechnical

ecosystems. Herein, we have illustrated that designing complex human-AI systems is better considered as a continuous process that produces dynamic, adaptive, and cooperative systems rather than fixed products. Surely, within the complex human-AI ecosystems of the future, all kinds of individual agents will exist. Among humans, there will be leaders and followers, generalists and specialists, untrained and trained, and so on. Likewise, among technologies, we expect to see everything from single-purpose robots to increasingly intelligent cognitive agents and beyond. Questions about how to select a group of humans and AIs to develop an effective team will require a dynamic process that involves similarly complex dynamics among the team of developers and end-users. It is within these complex dynamics that we anticipate we will find both the greatest challenges to our success, as well as novel opportunities that may not have been recognized without first bringing the system together and observing its collective capabilities.

References

[1] Harari YN. Homo Deus: a brief history of tomorrow. Vintage Popular science. London: Harvill Secker; 2016. OCLC: 962379989.

[2] Van Der Waa J, Verdult S, Van Den Bosch K, Van Diggelen J, Haije T, Van Der Stigchel B, et al. Moral Decision Making in Human-Agent Teams: Human Control and the Role of Explanations. Frontiers in Robotics and AI. 2021;8.

[3] Metcalfe JS, Perelman BS, Boothe DL, Mcdowell K. Systemic Oversimplification Limits the Potential for Human-AI Partnership. IEEE Access. 2021;9:70242-60.

[4] Mostafa SA, Ahmad MS, Mustapha A. Adjustable autonomy: a systematic literature review. Artificial Intelligence Review. 2019;51(2):149-86.

[5] Peeters MM, van Diggelen J, Van Den Bosch K, Bronkhorst A, Neerincx MA, Schraagen JM, et al. Hybrid collective intelligence in a human–AI society. AI & SOCIETY. 2021;36(1):217-38.

[6] Fitts PM, editor. Human engineering for an effective air-navigation and traffic-control system. Human engineering for an effective air-navigation and traffic-control system. Oxford, England: National Research Council, Div. of; 1951.

[7] Sheridan TB. Function allocation: algorithm, alchemy or apostasy? International Journal of Human-Computer Studies. 2000 Feb;52(2):203-16. Available from: `http://www.sciencedirect.com/science/article/pii/S1071581999902859`.

[8] Nothwang WD, McCourt MJ, Robinson RM, Burden SA, Curtis JW. The human should be part of the control loop? In: 2016 Resilience Week (RWS); 2016. p. 214-20.

[9] Norman DA, Draper SW. User centered system design: New perspectives on human-computer interaction. 1986.

[10] Pea RD. User centered system design: new perspectives on human-computer interaction. Journal educational computing research. 1987;3(1):129-34.

[11] Jiao P, Veiga A, Walther A. Social media, news media and the stock market. Journal of Economic Behavior & Organization. 2020;176:63-90.

[12] Marathe A, Brewer R, Kellihan B, Schaefer KE. Leveraging wearable technologies to improve test & evaluation of human-agent teams. Theoretical Issues in Ergonomics Science. 2020;21(4):397-417.

[13] Rasmussen J, Goodstein L. Decision support in supervisory control of high-risk industrial systems. Automatica. 1987;23(5):663-71.

[14] Kozlowski SW, Klein KJ. A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. 2000.

[15] Salas E, Stagl KC, Burke CS, Goodwin GF. Fostering team effectiveness in organizations: toward an integrative theoretical framework. 2007.

[16] Marks MA, Mathieu JE, Zaccaro SJ. A temporally based framework and taxonomy of team processes. Academy of management review. 2001;26(3):356-76.

[17] Malone TW. Superminds: the surprising power of people and computers thinking together. First edition ed. New York: Little, Brown and Company; 2018. OCLC: on1003309442.