Project Alloy – Machine Learning Challenges for Researching Human-Machine Teaming

Laura Seaman[1*], Tomo Lazovich[1], Israel Ridgley[1], Michelle Beard[1], Troy Lau[1], Megan Mitchell[1], Joshua C. Poore[1]

[1] Draper, 555 Technology Sq, Cambridge MA, 02139

[*] Corresponding author: lseaman@draper.com

Intelligent machines invoke complex mathematics to process volumes and varieties of data at speeds well beyond the capability of a human, providing the potential to transform nearly every human pursuit, from core scientific pursuits to fighting wars in the battlefield and cyberspace. However, current approaches to developing true collaborations within hybrid human–machine teams fail to invoke these intelligent machines as partners because: (1) team designers lack experimental platforms for testing performance and deployment strategies for new machine intelligence partners; (2) closed-loop team management lacks adequate sensing to observe the changing states of the individual team members and the evolving problem space; and (3) deployment strategies too narrowly scope the role of collaboration by limiting machine partners to fixed, bounded tasks, denying team leaders the ability to adapt to dynamic, multi-granular changes. By creating a machine learning challenge based testbed to evaluate performance, sense the states of team members and their environment, and precisely effect change in hybrid teams, we are eliminating these barriers to field-changing breakthroughs in the science of human-machine teaming.

Project Alloy (funded by DARPA's Agile Teams program) aims to develop and implement intelligent machine agents that team with humans (hybrid teams) in meaningful and supportive ways. Further, this project aims to leverage real-world web platforms as test beds to both prove out the effectiveness of these agents, but also identify gaps in their capabilities to support growth and technological readiness. The following three goals will allow us to study how to develop, evaluate and maximize the effectiveness of hybrid teams:

Goal 1: Machine learning challenges as an experimental platform. By hosting challenges in a mature citizen science competition platform, we provide a unique environment to systematically develop and test hybrid team theory and algorithms. Draper's 2016 "Satellite Image Chronology" contest with Kaggle pushed the boundary of the contest format, allowing for the first time human inputs to be added to data science solutions. We will introduce novel contest features by increasing the complexity of challenges to require larger, multidisciplinary teams, and gradually reveal additional project goals to stimulate adaptation. During the machine learning challenges a unique range of machine intelligent agents will be provided for use as team members matched with a powerful incentives structure for teams to maximize the usage of these partners. This approach is independent of the specific competition platform ultimately used, but its power is clear. For example, Kaggle has a community of more than 500,000, and a single contest will often attract over 1000 participants who are motivated by the opportunity to solve compelling real world problems, as well as the monetary prizes and glory of winning. Using machine learning challenges as an experimental platform has the benefit of enabling cost effective, large scale, reproducible hypothesis testing.

Seven machine learning competitions will be held over 4 years. The primary goals are to: 1) create challenges that will attract hundreds of teams; 2) necessitate multiple roles and agile,

adaptive development; and 3) develop technologies that will facilitate learning about human-machine interactions. Several challenge domains will be explored including autonomous vehicles, neuroscience, and cyber defense. Over seven challenges, we expect to attract over 6000 participants.

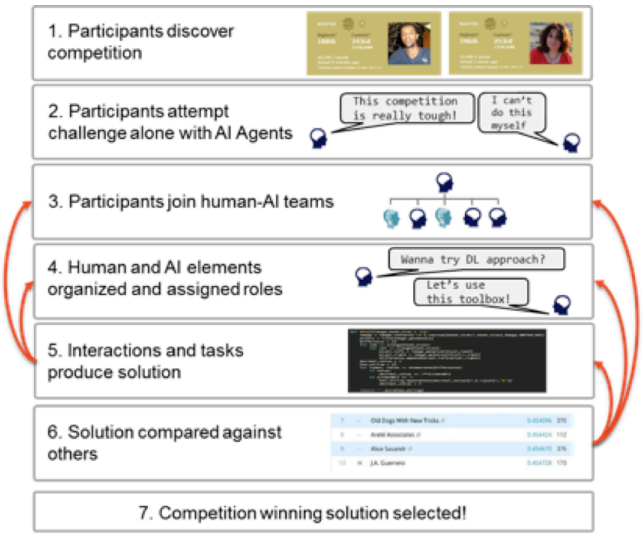

Figure 1 – Machine learning experiment example

Figure 1 shows how participants will experience an AI-infused machine learning challenge. Participants can attempt to solve the challenge alone, with AI partners, or with a fully hybrid team that iteratively progresses towards the goal. The key enabler in these experiments is an online, citizen science competition platform and a robust infrastructure for submission and scoring. As teams submit answers they are scored on independent servers and each new score is reflected on a leaderboard in real-time. Existing platforms enable members to nominate team leaders and support team mergers over the course of a challenge.

Goal 2: Workspace instrumentation. By providing dynamic sensing of each team's state and progress towards a solution, we will enable researchers to develop and test analytical formulations of hybrid team management. We will embed instrumentation into a code development environment to provide in-depth measurements of team performance. The combination of team performance monitoring and leaderboard scoring transforms the contest platform into a world-wide social science and machine intelligence laboratory.

Goal 3: Machine intelligent partners for problem solving and team building. By providing machine agents that can augment or substitute human roles, we can explore a tighter synthesis of human and machine strengths for greater resilience and agility under changing project goals and constraints. Current practices of substituting machine intelligences for fixed, bounded tasks work well with projects that progress neatly though planning and execution phases. Agile teams must rapidly shift and be able to continuously monitor goals and constraints, evaluate alternatives, test potential strategies, and redeploy resources. Some provided agents assess the team and project goals to aid with task breakdown and role assignment while other agents can analyze and execute software engineering tasks.

During the workshop discussion, we are interested in learning:

- What types of challenges garner the most interest and desire to create teams?
- How should incentives be structured to account for multiple types of teams?
- How do we make participants feel like partners in the research?

The challenge problems produced will serve as leave behinds to motivate future research in human-machine teaming. Because our challenges are software development problems for machine learning, the citizen scientist responses will provide new solutions and strategies in machine learning. Project Alloy will not only support evaluation of algorithms, but also provide legacy resources that motivate future hybrid team research and transition. This work is partially sponsored by the Air Force Research Laboratory (AFRL) and DARPA.