Get Together in the Middle-earth: a First Step Towards Hybrid Intelligence Systems

GIOVANNA VARNI, LTCI, Télécom Paris, Institut polytechnique de Paris, France
ANDRÉ-MARIE PEZ, LTCI, Télécom Paris, Institut polytechnique de Paris, France
MAURIZIO MANCINI, Sapienza University of Rome, Italy

In the last decade, the number of computer systems using AI has increased dramatically. To date, indeed, AI is present in almost all the aspects of the human everyday life. This resulted in the attempt of scholars in Computer Science to endow machines with human-like socio-cognitive skills and/or human-like embodiment to try to improve interactions. Such an approach, however, highlights several crucial issues related to the substantial differences between fine-grained human skills and what machines can do and learn. So, although being expensive and sophisticated tools, machines tend to be "idiots savants". Hybrid Intelligence (HI) is aimed to tackle this issue by proposing, as Akata and colleagues say, "systems that operate as mixed teams, where humans and machines cooperate synergistically, proactively, and purposefully to achieve shared goals". To our knowledge, however, HI is at a very early exploratory stage, and few concrete solutions to deal with it exist. In this position paper we introduce and briefly describe "Middle-Earth", a conceptual and experimental ground to study HI. Moreover, we present a first prototype of a software platform based on immersive VR environments, on which we plan to carry out in the future the first pioneering experiments on teams of humans and/or AI-driven agents getting together in Middle Earth to perform collaborative tasks.

CCS Concepts: • **Human-centered computing** → *Human computer interaction (HCI)*; **Computer supported cooperative work**; *Virtual reality*.

Additional Key Words and Phrases: hybrid intelligence, collaborative, team, teammate, AI

ACM Reference Format:
Giovanna Varni, André-Marie Pez, and Maurizio Mancini. 2021. Get Together in the Middle-earth: a First Step Towards Hybrid Intelligence Systems. In *Companion Publication of the 2021 International Conference on Multimodal Interaction (ICMI '21 Companion), October 18–22, 2021, Montréal, QC, Canada.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3461615.3485413

1 INTRODUCTION

Despite the increasing pervasiveness of AI in computer systems, such systems mainly exploiting abstract intelligence, tend to be "idiots savants", surpassing the performance of human experts in a very narrow range such as performing calculations [1]. Conversely, humans intelligence has a number of different and complementary facets, enabling people, for example, to get along with others, to create artifacts or thoughts, that are original, valuable and unexpected. Human-Computer-Interaction (HCI) traditionally studies paradigms for improving the interactions between humans and computers. HCI endeavours to make computers more usable, safe and close to the users' needs. Currently, a convergence of HCI and AI is being observed to design computers that are both interactive and intelligent. A concrete outcome of this process is the ACM Transactions on Interactive Intelligent Systems (TiiS) [1]. More recently, a new research field, called Hybrid Intelligence (HI), emerged. This

[1]https://dl.acm.org/journal/tiis

ICMI '21 Companion, October 18–22, 2021, Montréal, QC, Canada
© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.
This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *Companion Publication of the 2021 International Conference on Multimodal Interaction (ICMI '21 Companion), October 18–22, 2021, Montréal, QC, Canada*, https://doi.org/10.1145/3461615.3485413.

aims at "creating systems that operate as mixed teams, where humans and machines cooperate synergistically, proactively, and purposefully to achieve shared goals" [1]. In this view, the role played by machines shifts from being very sophisticated but somewhat passive tools to becoming proactive teammates. Scholars identify four main properties of HI, that also define four research lines that should be investigated in the near future [1]: (1) *Collaborative*, to promote synergy between humans and machines; (2) *Adaptative*, to allow computers learning from and adapting to humans in their environment; (3) *Responsible*, to ensure that machines show ethical behavior, by avoiding, for example, that they will "rule the world"; (4) *Explainable*, to develop a shared workflow, from reasoning to acting. Research is ongoing in these directions, to effectively achieving HI by leveraging the peculiar skills of humans together with those of AI. Therefore, the concept of HI overcomes the ongoing idea of building increasingly human-like machines, by exploiting AI to amplify, instead of replacing, human intelligence. While HI is a revolutionary and intriguing concept, related research is at an early and exploratory stage, as witnessed by the recent "Research Agenda for Hybrid Intelligence" [1] and the "Research Agenda on AI in Team Collaboration" [14].

In this view, this position paper provides a twofold contribution: i) presenting *Middle-earth*, a conceptual and experimental ground to study and explore HI; ii) describing a first prototype of a flexible software platform which in the future may support a concrete realization of this ground. More specifically, we propose to study HI through immersive VR environments. We strongly believe, indeed, that VR environments could become the Middle-earth in which humans and machines get together to synergistically "achieve goals that were unreachable by either humans or machines alone" [1], in such a way that, as stated by Aristotle, "the whole is greater than the sum of its parts".

We plan to address HI through immersive VR environments (i.e., the Middle-earth) by following the four lines of research mentioned above and in [1]. More in detail, in this paper we discuss the first one, that is, the *Collaborative* property. We believe, indeed, that, among the four research lines, this is the one from which the other three stem from. However, Middle-earth is suitable for investigating all the facets of HI. As mentioned in the documentation available in the website of the "Hybrid Intelligent Centre" [2], the Collaborative property open research questions include:

- How can machines adapt to the ways humans typically collaborate?
- How can machines amplify human capacity to collaborate effectively?
- Can we build machines that can learn what these are, adapt over time via social learning (similar to humans)?
- Could machines become part of groups and have a positive influence on people's willingness and competence to cooperate, and so producing better outcomes for the group as a whole?
- Can machines learn to recognise social practices and collaborative strategies of humans?
- Can machines contribute to the learning, setting and execution of work agreements?

2 THE 3 PILLARS OF MIDDLE-EARTH

The investigation along the Collaborative research direction in Middle-earth is supported by the three main pillars depicted in Figure 1:

- *Introducing hybrid sensing* - Despite human senses are more refined and balanced from a qualitative point of view, the machine ones (that are usually made to try to mimic the human ones) provide more quantitative and precise information. Let us consider, for example, a piece that needs to be joined to another one using a screw. There is a box containing several screws, having different pitches. The box also contains other stuff such as bolts, nails, hooks, and so on. A human can promptly distinguish screws from the other tools, whereas a machine can easily quantify the pitch of the screw. The joint work of the human and the

[2] https://www.hybrid-intelligence-centre.nl/

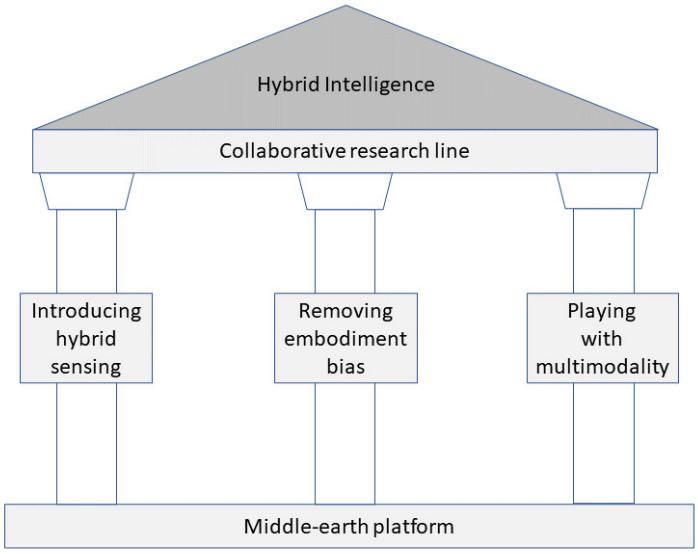

Fig. 1. The 3 Pillars of Middle-earth: 1) Introducing hybrid sensing, 2) Removing embodiment bias, 3) Playing with multimodality.

machine can thus lead to an immediate and robust identification of the screw that best suits the purpose. By allowing the combination of both qualitative and quantitative information, Middle-earth will enable the emergence of an hybrid sensing for solving tasks that a human or a machine alone could not effectively perform without a specialized training.

- *Removing embodiment bias* - As defined in [7], Embodiment is "the property of being manifest in and of the everyday world" and it "denotes not physical reality but participative status". In Middle-earth, humans and machines can share the same level of embodiment, that is, they cannot be distinguished neither from their physical appearance nor from their participative status. This will allow researchers to study interaction without worrying about human *a prioris* about machines. This is a very crucial point for achieving an effective HI, in fact, de Graaf and Malle argue that "the anthropomorphization of agents causes users to expect explanations utilizing the same conceptual framework used to explain human behaviors" [6]. As human and machine embodiment can look identical in Middle-earth, human trust in the machine can be increased. Previous work shows, indeed, that humans tend to project their social expectations and norms on anthropomorphic machines, like they would do with other humans [9], and trust is lost when humans cannot understand observed behaviors or decisions [1].
- *Playing with multimodality* - Existing studies show the importance of multimodal interaction for collaboration [11]. Humans are intrinsically multimodal, while reaching multimodality for a machine is not a trivial task, as demonstrated by the open challenges in the field of multimodal machine learning [2], such as: multimodal representation of information, alignment between modalities, and fusion of modalities. In Middle-earth, communication modalities can be easily manipulated, both on the human and the machine side, by inhibiting and/or enhancing them, to study how they are understood for the purpose of collaboration. So, it will enable researchers to explore in detail multimodality and to propose new design spaces for collaborative multimodal systems.

3 THE MIDDLE-EARTH PLATFORM PROTOTYPE

This section presents a first prototype of the Middle-earth platform, which will allow human and/or AI-driven teammates to interact through several modalities by sharing the same virtual environment and the same embodiment, as described in Section 1. This prototype exploits the auditory and the visual modalities. We plan to adopt this prototype to conduct initial studies on HI. The platform has three main components:

- *server* - it manages the message exchange between instances of the other two components; whenever a new client (player or master) component connects for the first time, the server assigns a unique id to the client, sending a message containing the id of the current scene and the list of the players;
- *player (client)* - it allows the user, that is, a human or a AI-driven agent, to see and interact with the virtual environment; for example, they can catch or activate objects (e.g., turning on and off the light, opening and closing windows and doors, and so on); in the current implementation, the player component has to be run on a machine equipped with a HTC Vive headset, allowing human users to interact with the other teammates from any location, being local or remote;
- *master (client)* - through this client component, the experimenters can observe the players' interaction, trigger the currently played scene, start and stop a timer that will be seen by the players and talk to the players through a microphone. In the future, the experimenters can also change the type of the users' embodiment.

3.1 Audio/Video Rendering

After the initial connection to the server, each player component sends, in real-time, the coordinates of the user's hands, head and upper body. These coordinates are extracted by the HTC Vive headset. Also, the player component streams the audio of the microphone to the server. In turn, the server component broadcasts this data to the other connected components, being players or masters. Finally, each player component displays the user's hands in the user's field of view, so the user can see them in the virtual environment.

To avoid motion sickness, which can affect users observing the scene rendered through the VR headset, all the scene changes and teleportation actions are preceded and followed by a fading effect (i.e., the scene is faded to/from back).

3.2 Object Operation and Activation

Each scene contains a number of objects that can be operated or activated/deactivated by the players. When getting close enough to one of these objects, a ray of light is cast from the user's (virtual) hand toward the object, on which a small sphere of light is also rendered, to indicate that the object can be grasped. Three types of interaction are available:

- *teleportation* - users can teleport themselves in the locations highlighted by the ray of light;
- *catch* - user can catch an object if it is closer than 30 centimeters from the their hand, to examine it, and then release it; only one user at a time can catch an object, with only one hand at a time; the actions of catching and releasing objects are broadcast by the server to all the player and master components, so they can play back the corresponding animations on the client headsets;
- *activation* - some objects can have multiple states (e.g., open/closed, on/off); whenever a user activates or deactivates one of these objects, its state changes and the corresponding animation is played back (some animations will be played in a loop, e.g., when the fan is turned on); these state changes are broadcast by the server to all player and master components, both the already connected and the newly connected ones.

3.3 The Master's Special Powers

The experimenter playing the role of the master is responsible of managing the evolution of the scenario currently played by the users. There can be more than one master at the same time. The master's speech is broadcast through the server to all the connected components in real-time. Also, the master can trigger the scene currently played by the connected users. For example, the initial scene will be triggered and maintained until all the users are connected; when that happens, the master will announce that the scenario is starting and will trigger the next scene. The master user can start and stop a timer, that is broadcast to all the connected player components to be displayed locally to the users. Depending on the current scene, the timer will allow the master to trigger events when, for example, the time reaches zero.

3.4 Messages

The server and (player and master) client components communicate through network messages that are sent via UDP (to keep the time needed to transmit a message at the minimum). Each message is labelled with the id of the sender: 0 indicates that the sender is the server, while any id greater than 0 identifies a player (the id is assigned by the server to each client at the time of the first connection). The messages that are exchanged between the server and clients are:

- *position/rotation*: containing the real-time position and rotation of the objects in the scene and the users' heads, torsos and hands;
- *audio*: containing audio buffers coming from the microphone of one of the players or master(s);
- *catch*: indicating the id of the player that is catching or releasing one of the objects in the environment;
- *activate/deactivate*: referring to the id of an object in the environment that is activated or deactivated;
- *scene*: indicating that a particular scene has to be loaded and started by the server, players and master(s);
- *timer*: the status of the master's timer.

4 STUDYING COHESION IN MIDDLE-EARTH

When more than two people interact together in joint activities, social processes resulting from such affective, behavioral and cognitive interactions may arise. These processes are called "emergent states" [10] and influence the team outcomes, such as effectiveness and performance [13]. Current research on emergent states almost exclusively deals with teams composed by humans eventually supported by technology (e.g., communication technology). In [5] the way in which some technological tools impact emergent states as team cognition, team trust, cohesion and conflict is investigated. To the best of our knowledge, however, the specific role of AI on the development of emergence states is under-investigated. To illustrate how Middle Earth could be adopted for future research, we provide an example about team cohesion that is one of the most investigated emergent state. Cohesion can be defined as "a dynamic process that is reflected in the tendency for a group to stick together and remain united in the pursuit of its instrumental objectives and/or for the satisfaction of member affective needs" [4], and it is considered as a highly valued group property serving crucial roles for group effectiveness and performance. We believe that through Middle Earth the following issues can be investigated:

- how cohesion manifests in HI scenarios where there are a variable number of human and/or AI-driven agents acting together;
- whether and how theories and models of cohesion holds in HI scenarios;
- which modalities and which combinations of them are the most suitable to enhance collaboration in HI scenarios.

As an example, we describe here an experimental scenario that could be used to address the points listed above. The scenario is an escape room game. Social games, such as escape rooms, are a form of socially rich multi-party interaction where people coordinate and collaborate to achieve common goals. They are a viable methodology to

Fig. 2. Players collaborating in the Middle-earth escape room scenario.

catch the subtle nuances of human-human communication in several research domains, from psychology [8], neuroscience [12] and human computer interaction [3]. Figure 2 shows players in the escape room scenario.

In particular, we conceived the following 3 tasks, each one addressing a different level of cohesion:

- *task 1: secret code (individual)* - participants are teleported by the master into a room that they have to explore to find the digits of the secret code that, in task 2, will be used to open a locked briefcase; the digits are written on some playing cards hidden inside or behind some objects in the room; a 5 minutes timer is displayed by each user, adding pressure to the task;
- *task 2: briefcase opening (collaborative, cognitive)* - participants are teleported into the same room of task 1, and have 5 minutes to share the information they found about the secret code, and agree on the final code; they can try to open the briefcase 4 times max;
- *task 3: unknown object (collaborative, operational, creative)* - participants are teleported again into the room, in which the briefcase is now open (even if they could not manage to open it in task 2), disclosing 3 unknown objects; they now have 3 minutes to come up with a story linking the 3 objects together.

5 CONCLUSION

Hybrid Intelligence (HI) is a revolutionary and intriguing concept, aiming at "creating systems that operate as mixed teams, where humans and machines cooperate synergistically, proactively, and purposefully to achieve shared goals" [1]. This position paper introduces Middle Earth a conceptual and experimental ground to allow researchers designing and carrying out pioneering investigations on HI. A first prototype of a software platform which may support a concrete realization of this ground is also presented.

6 ACKNOWLEDGMENTS

This paper has been partially supported by the French National Research Agency (ANR) in the framework of its JCJC program (GRACE, project ANR-18-CE33-0003-01, funded under the Artificial Intelligence Plan).

REFERENCES

[1] Zeynep Akata, Dan Balliet, Maarten De Rijke, Frank Dignum, Virginia Dignum, Guszti Eiben, Antske Fokkens, Davide Grossi, Koen Hindriks, Holger Hoos, et al. 2020. A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect With Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer* 53, 8 (2020), 18–28.

[2] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2018. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence* 41, 2 (2018), 423–443.

[3] C. Bonillo, T. Romão, and E. Cerezo. 2019. Persuasive games in Interactive Spaces: The Hidden Treasure Game. In *Proceedings of the XX International Conference on Human Computer Interaction*. 1–8.

[4] Albert V Carron and Lawrence R Brawley. 2000. Cohesion: Conceptual and measurement issues. *Small group research* 31, 1 (2000), 89–106.

[5] Petru Lucian Curşeu. 2006. Emergent States in Virtual Teams: A Complex Adaptive Systems Perspective. *Journal of Information Technology* 21, 4 (2006), 249–261.

[6] Maartje de Graaf and Bertram Malle. 2017. How People Explain Action (and Autonomous Intelligent Systems Should Too). In *AAAI Fall Symposium Series*. 19–26.

[7] Paul Dourish. 1999. Embodied interaction: Exploring the foundations of a new approach to HCI. *Unpublished paper, on-line: http://www. ics. uci. edu/˜ jpd/publications/misc/embodied. pdf* (1999).

[8] G. Freedman and M. Flanagan. 2017. From dictators to avatars: Furthering social and personality psychology through game methods. *Social and personality psychology compass* 11, 12 (2017), e12368.

[9] Zahra Rezaei Khavas, Reza Ahmadzadeh, and Paul Robinette. 2020. Modeling Trust in Human-Robot Interaction: A Survey. arXiv:2011.04796 [cs.RO]

[10] S. W. J. Kozlowski. 2015. Advancing research on team process dynamics: Theoretical, methodological, and measurement considerations. *Organizational Psychology Review* 5, 4 (2015), 270–299.

[11] Ingmar Rauschert, Pyush Agrawal, Rajeev Sharma, Sven Fuhrmann, Isaac Brewer, and Alan MacEachren. 2002. Designing a human-centered, multimodal GIS interface to support emergency management. In *Proceedings of the 10th ACM international symposium on Advances in geographic information systems*. 119–124.

[12] E. Redcay and L. Schilbach. 2019. Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience* 20, 8 (2019), 495–505.

[13] R. Rico, M. Sánchez-Manzanares, F. Gil, and C. Gibson. 2008. Team Implicit Coordination Processes: A Team Knowledge-Based Approach. *The Academy of Management Review* 33, 1 (2008), 163–184.

[14] Isabella Seeber, Eva Bittner, Robert O Briggs, Triparna De Vreede, Gert-Jan De Vreede, Aaron Elkins, Ronald Maier, Alexander B Merz, Sarah Oeste-Reiß, Nils Randrup, et al. 2020. Machines as teammates: A research agenda on AI in team collaboration. *Information & management* 57, 2 (2020), 103174.