Revealing and Mitigating Harmful Assumptions and Behaviors in Human-Autonomy Teaming

by

Christine T. Chang

B.S., Cornell University, 2005

M.S., Boise State University, 2013

M.S., University of Colorado Boulder, 2021

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Computer Science

2024

Committee Members:

Bradley Hayes, Chair

Clayton H. Lewis

Nisar Ahmed

Casey Fiesler

Tristan Endsley

Chang, Christine T. (Ph. D., Computer Science)

Revealing and Mitigating Harmful Assumptions and Behaviors in Human-Autonomy Teaming

Thesis directed by Prof. Bradley Hayes

Human-autonomy teaming in complex environments continues to evolve with technological innovations like mixed reality and rapidly improving large language models. With this evolution comes a need for increased safety measures and better ways for humans to learn and understand these systems. The work presented in this dissertation aims to address questions about safety, appropriate trust, and appropriate use of autonomy by and for humans. I begin with an overview of how mixed reality, and mainly augmented reality, is used for human-robot collaboration. I then explore how we might use augmented reality to promote safety and compliance in a shared space environment with humans and robots. This leads to the question of how we can actively warn humans about failures of autonomous chatbots. And finally I investigate the use of iteratively adding latent human knowledge to an autonomous robot's trajectory optimization as a way of improving both learning and mission outcomes. Ultimately I show that humans have a propensity to dangerously overtrust robots and other forms of autonomy, however we can mitigate this bias with certain design considerations including iteration and transparency.

Dedication

For Mike Eisenberg, who took a chance on me when I started this adventure,

for Carolina Restrepo, whose life and legacy will continue to inspire me,

and for Chuck Burnell, whose example of how to live in kindness pushes me daily.

These beautiful and brilliant people all left us too soon.

Acknowledgements

I am sincerely grateful to my advisor and committee chair, Dr. Bradley Hayes. This work would not have been possible without his guidance and support, including his willingness to bring a true novice into his lab. I also thank my Draper mentor Mitchell Hebert for always enthusiastically lending me an ear and the Draper Scholar Program for generously funding my research.

I am deeply thankful to my lab mates and co-authors, from both the Craft Tech Lab and the CAIRO Lab, as well as to my friends from across the department, campus, and Boulder community who made my time as a Ph.D. student immensely more enjoyable. The lunches, coffees, accountabilibuddy pomodoros, trail runs and bike rides, happy hours, Zoom hangouts, and times we've said "We should start a podcast" will not soon be forgotten. To persist in a doctoral program through the loss of an advisor, a global pandemic, a mass shooting, and deaths of close friends takes the support of many people. You all know who you are, and I am so glad to call you my friends.

And finally, thank you to my partner, Jason, who picks up all my slack, and to my parents, who raised me to think that I could do anything. You have all always believed in me, even when I wasn't sure I believed in myself. I am proud to be the person I am today. I am also proud to be Astro's cat mom – his belly rubs are the best.

Contents

Tables

Table

Figures

Figure

Chapter 1

Introduction

1.1 Motivation

As the prevalence of human-autonomy teams continues to increase, so do the potential benefits as well as the risks. These autonomous systems range from physical robots to automated decision-making systems to chatbots and everything between. However, these systems are frequently confusing, misleading, or mysterious, leading to misuse, overtrust, and failure to achieve the team objectives.

In 1951 the famous Fitts list [1] claimed that humans are better at certain activities while machines are better at others (often referred to as "MABA-MABA" for "men are better at/machines are better at" or "HABA-MABA" for "humans"). While this remains true to an extent, the lines are rapidly becoming blurred and crossed. Furthermore, much of the current HRI, HCI, and AI literature could form a convincing argument that the combination of humans and "machines" (or autonomy in general, whether embodied in a robot or not) can have a stronger impact than just the sum of their parts. Measures of human-autonomy team fluency [2] can elucidate and quantify that combination. However, key requirements for such fluency include safety, understanding, situational awareness, transparency, and explainability.

While physical robots, computer-based chatbots, and other autonomous systems inhabit different physical spaces in our world, the dangers and benefits of both have significant overlap. The work presented in the pages that follow seeks to clarify and explain autonomous systems for human teaming. It also aims to uncover insights that can be used to directly inform designers and users of these systems.

1.2 Thesis Statement

With the work presented in this thesis, I surface insights about human behavior when operating with and around autonomous systems, and I show that there are specific design principles by which human-autonomy teaming systems can intentionally mitigate harmful human assumptions and behaviours, achieve increased human understanding, and improve overall team outcomes.

1.3 Overview

This thesis begins with an in-depth discussion of the use of augmented reality and some other forms of extended reality for human-robot collaboration in Chapter 2. We explore how these technologies have been developed over the past few decades and different uses for them with human-robot teams. This literature review led me to ask whether one can use this rapidly expanding innovation to help keep people safe in collocated situations with robots.

In the first study that I present in Chapter 3, we develop a system that places a human and an autonomous robotic quadcopter in a shared space environment [3]. Equipped with an augmented reality head-mounted display, the human can communicate with the robotic system regarding ownership of the space in a warehouse scenario. We learn that people sometimes fail to comply with instructions given by an autonomous system, even in this potentially dangerous situation with a close-proximity quadcopter. The results of this study begged the question of whether we can convince people to proceed with more caution when given explicit warnings.

To test this question, in Chapter 4 we designed chatbots driven by large language models [4]. One chatbot was intended to support participants writing a short essay and another chatbot was developed as a resource for a bridge design task. We then embedded false information into some of the chatbot responses and equipped both chatbots with specific kinds of warnings. Through this study we learned that people largely disregard any explicit warnings in chatbots, even when those warnings are shown in a variety of ways – including ways that are commonly used in chatbots today.

Still in search of an appropriate method for appropriately mediating humans' interactions with autonomous systems, in Chapter 5 we asked how we could compel people collaborating with autonomous systems to proceed slowly and deliberately [5]. To test this, we compared outcomes between an interface that allowed a human in the loop to provide latent knowledge to a trajectory optimization iteratively and an interface that required all latent human knowledge in a single iteration. We learned that an iterative affordance produces better outcomes for the system, the mission, and the human learner, supporting the user in a similar way to scaffolding in learning.

I conclude in Chapter 6 by examining how can we encourage responsible use of the technologies explored in this dissertation. I set some concrete directions for future work and frame some recommendations for public policymakers.

Chapter 2

Extended Reality for Human-Robot Collaboration

2.1 Introduction

Augmented reality (AR) has been explored as a tool for human-robot collaboration (HRC) since 1993 in [6], and research related to AR for HRC has expanded further with the deployment of the Magic Leap 1 [7] and Microsoft HoloLens 2 [8], arguably the most advanced head-mounted displays for AR on the market. In 2008, Green, Billinghurst, Chen, **et al.** [9] presented a literature review of AR for human-robot collaboration, and in the years that have passed since then, AR for HRC has evolved immensely. The ACM/IEEE International Conference on Human-Robot Interaction hosts annual workshops on Virtual, Augmented, and Mixed Reality for Human-Robot Interaction (VAM-HRI) [10]–[14], further evidence that these technologies of augmented reality and robotics are becoming increasingly used together. Virtual, augmented, and mixed reality are often classified together as "Extended Reality" or XR. This survey is largely intended to be a continuation and expansion of the review begun by Green, Billinghurst, Chen, **et al.** [9].

Milgram, Zhai, Drascic, **et al.** [6] define augmented reality as an overlay of virtual graphics and virtual objects within the real world, and this is the basic definition used throughout this paper. Green et al. add that "AR will allow the human and robot to ground their mutual understanding and intentions through the visual channel affording a person the ability to see what a robot sees" [9]. Whether the real world is viewed unobstructed, partially obstructed, or through an intermediate display, the AR features are placed over these real world images. Technologies that enable augmented reality include mobile devices such as head-mounted displays or handheld tablets, projection-based displays, and static screen-based displays, and are detailed in Section 2.2. This paper aims to focus on the topics of *augmented reality* as applied specifically to *human-robot collaboration*, and thus *excludes* related but different topics such as virtual reality, augmented virtuality, or augmented reality for purposes other than HRC. Because human-robot collaboration occurs across all types of robots, we include examples of this variety within every section.

We conducted this literature review by searching the proceedings of highly-refereed robotics, human-robot interaction, and mixed-reality conferences, as well as associated journals. Conference proceedings and journals included the ACM/IEEE International Conference on Human-Robot Interaction (HRI), Robotics: Science and Systems (RSS), International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS), IEEE International Conference on Robot and Human Interactive Communication (ROMAN), IEEE International Conference on Intelligent Robots and Systems (IROS), IEEE International Conference on Robotics and Automation (ICRA), ACM/IEEE Virtual Reality International Conference (IEEE VR), IEEE International Conference on Control, Automation, and Robotics (ICCAR), IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), CIRP Annals: Journal of the International Academy for Production Engineering, IEEE International Conference on Mechatronics and Machine Vision in Practice (M2VIP), IISE Transactions, Transactions on HRI, Frontiers in Robotics and AI, Frontiers in VR, and ICAR. Keywords utilized for the search were "augmented reality" and "mixed reality". If the conference or journal was not robotics focused, the keyword "robot" was also used. We recognize that this method does not elicit a fully comprehensive review of all literature on HRC via AR, however we believe that our sample size is large enough to be representative of where the field has been and is heading.

We then examined the literature around augmented reality for human-robot collaboration, using the following questions to determine how to organize the discussion for each article:

- Is the contribution primarily about helping to program, create, and/or understand a robot and/or system?

- Is the contribution primarily about improving the collaborative aspects of a human-robot interaction?

In many cases there is significant overlap in these contributions and thus multiple valid possible organizations of these works. For this article we use the more significant area of contribution to situate the research with respect to other relevant literature.

First we begin by exploring the many different manifestations of AR as it has been used for HRC since 2008 (Section 2.2). We then highlight the literature as it represents the categories defined above in Sections 2.3 and 2.4. Section 2.5 reviews a representative selection of the evaluation strategies and methods utilized in the related studies. Section XXXXX provides a potential method of classifying XR for HRI, as proposed in [15]. And we conclude with a vision for where research on AR for HRC might be most useful in the future, including in space exploration applications[16] (Section 2.7).

Contributions and Categorizations of Included Papers

Modalities	
Mobile Devices: Head-Mounted Display	[17]–[34]
Mobile Devices: Handheld Display	[35]–[43]
Projection-based Display	[44]–[47]
Static Screen-based Display	[48]–[51]
Alternate Interfaces	[52]–[55]
AR Combinations and Comparisons	[46], [56]–[59]
Creating and Understanding the System	
Intent Communication	[43], [47], [57], [60]–[66]
Path and Motion Visualization and Programming	[18], [21], [32], [35]–[37], [39], [44], [46], [51], [52], [58], [62], [67]–[81]
Adding Markers to the Environment	[21], [35], [40], [51], [82]–[84]
Manufacturing and Assembly	[24], [25], [38], [44], [46], [67], [73], [84]–[86]
Improving the Collaboration	
AR for Teleoperation	[20], [23], [25], [29], [33], [48], [50], [56], [87]–[90]
Pick-and-Place	[28], [40], [51], [57], [58], [91]
Search and Rescue	[31], [55], [62], [69], [92]–[96]
Medical	[30], [34], [48], [97]–[101]
Space	[102], [103]
Safety and Ownership of Space	[3], [40], [41], [47], [73], [86]
Other Applications	[104]–[108]

Table 2.1: This table summarizes the categories outlined in this literature review and lists the articles associated with each category. Many papers are cited multiple times.

2.2 Reality Augmented in Many Forms

Augmented reality can manifest in different forms, as some modalities are better suited for certain uses than others, and AR has evolved significantly in the last decade. Head-mounted displays are some of the most commonly considered AR devices, frequently used in cases where the person is collocated with a robot and needs the use of both of their hands. Mobile phones and tablets offer a different experience with augmenting the real world, especially useful when those devices' other capabilities or apps might be utilized or to conduct smaller-scale interactions that do not necessitate an immersive view. Projection-based displays can be ideal for tabletop collaborative work or in consistent manufacturing environments, while static screen displays might best serve remotely located users. Below we discuss various modalities of AR, their uses, and how they have changed over time, particularly as applied to human-robot collaboration. We do this by presenting a list of works separated by AR modality due to the different interactions enabled and required. We also acknowledge that it is important to recognize the degree of abstraction of the robot platform communication as coming from a separate device.

2.2.1 Mobile Devices: Head-Mounted Display

Head-mounted displays (HMDs) for AR have increased in popularity for use in HRC as the technology has matured. Furthermore, since 2009 the research has evolved from showing basic prototypes and designs for using HMDs, as in Chestnutt, Nishiwaki, Kuffner, **et al.** [17], to more recently providing detailed design frameworks [18] and conducting extensive user studies with HMDs [19], [20], [27], [34].

Generally HMDs are used for in situ interactions with robots, whether aerial, tabletop, or ground-based. This way the virtual images (objects and/or information) can be placed over the physical objects within the environment that the user is currently experiencing. Depending on the maturity of the technology and the desired implementation virtual images can be either *egocentric* or *exocentric*. A helpful way to understand the difference between these two display types is to imagine a path being visualized. An exocentric display provides an external perspective of the path, such as a map, whereas an egocentric display provides a perspective from the point of view of a person actually traveling along that path. In the remainder of this subsection, we highlight literature that exemplifies the evolution of HMDs over time, while also indicating the multitude of ways in which they can be used to facilitate HRC.

In Chestnutt, Nishiwaki, Kuffner, **et al.** [17], the human user draws a guide path for a humanoid robot in the HMD, and the specific left and right footsteps are then shown to the user in their HMD such that they can anticipate where the robot will step. The robot plans its specific steps (shown as virtual footprints) based on the general path provided by the human (shown as a line drawing). In this paper written in 2009, all of these technologies are obviously still relatively nascent, a full user study is not conducted, and some alternatives to drawing the robot path are considered, such as joystick control. We see this change with modern research showing an increased expectation of rigor, a positive indicator of the field maturing.

Also in 2009, Green, Chase, Chen, **et al.** [21] utilize an HMD to allow a user to view virtual obstacles and plan a path for a simulated robot in AR. The HMD device used in the study, the eMagin Z800, was wired to a computer, and the work was done in simulation. This simulation-based work is further evidence of earlier studies finding ways to conduct AR-HRC research with still-maturing platforms.

Four years later in 2013, Oyama, Shiroma, Niwa, **et al.** [22] debut a "slenderized HMD"to provide a teleoperator the perspective of the robot. The device utilizes the same base HMD as in Green, Chase, Chen, **et al.** [21], but then also augments it with stereo cameras and a wide field of view camera. Similarly, the HMD in Krückel, Nolden, Ferrein, **et al.** [23] allows for teleoperation of an unmanned guided vehicle, but in this case the operator's view is augmented with an artificial horizon indicator and heading information. Furthermore, the operator can look around the entire environment, as they are effectively immersed in it with the use of the Oculus Rift HMD, a device intended for virtual reality more than augmented reality. This begs the question of what actually "counts" as AR; in the cases of Oyama, Shiroma, Niwa, **et al.** [22] and Krückel, Nolden, Ferrein, **et**

al. [23], the human's reality is not actually being augmented, they are instead being placed virtually into the environment of the robot. We claim that it is in fact augmented reality, since it is not a virtual environment that is being augmented. Despite the human not existing in the same location as the robot that they are controlling, a real environment is being augmented with virtual images, all of which the human user is able to see and affect.

The Microsoft HoloLens was introduced in 2016, facilitating a flurry of new research on AR for HRC using HMDs. Readers may note that the HoloLens is referenced throughout the literature mentioned in this paper, as it is relatively straightforward to work with and represents the state-of-the-art in augmented reality technology for head-mounted devices. The HoloLens 1 places images as holograms, or virtual images overlaid on the real world, in the wearer's field of view. This capability along with the incorporation of sensors allowing for detection of gaze, voice, and gesture made the HoloLens a revolutionary hardware development. In late 2019, the second version was released, HoloLens 2, with additional features and improvements including a more comfortable fit and eye tracking. The HoloLens has been mass produced for approximately 5 years now, making it widely available for research.

In Guhl, Tung, and Kruger [24], Guhl et al. provide a basic architecture for utilizing the HoloLens for industrial applications. Using tools such as Unity and Vuforia, robots can be modeled on the HoloLens, safety planes can be rendered to keep the human and robot safely separate, and sound can be played. These concepts and capabilities are suggested in hopes of allowing users to foresee robots' motions and thereby productively interfere.

Technology in Yew, Ong, and Nee [25] takes the AR user's environment and "transforms" it into the remote environment of the teleoperated robot. Real objects in the user's environment are combined with virtual objects in AR, such as the robot and the objects with which it is interacting, thereby reconstructing the actual site of the robot for the teleoperator.

A robotic wheelchair user in Zolotas, Elsdon, and Demiris [26] is outfitted with a Microsoft HoloLens. A rear-view display is provided, the future paths of the wheelchair are projected onto the floor, possible obstacle collisions are highlighted, and vector arrows (showing both direction and magnitude) change with the user-provided joystick velocity commands. One set of findings from this study was its deeper understanding of users' comfort with AR feedback. They also further confirmed the restrictive field of view of the HoloLens and cited it as a limiting factor in the usefulness of the AR. Work in Zolotas and Demiris [19] then builds on Zolotas, Elsdon, and Demiris [26] by adding "Explainable Shared Control" to the HMD. In this way the researchers aim to make the robotic wheelchair's reasoning more transparent to the user. The AR is classified as "environmental" (exocentric) or "embodied" (egocentric), depending on whether it is fixed to the environment or fixed to the user or robot. In another recent robotic wheelchair study using the HoloLens Chacón-Quesada and Demiris [27] test different types of icons and display modes. The user can control the wheelchair from within the AR interface, and a choice of movement options is shown to the user in their field of view.

The HoloLens was also used to program a UR5 robot arm to conduct pick and place tasks in Rudorfer, Guhl, Hoffmann, **et al.** [28]. The platform uses the built-in recognized HoloLens gestures to interact with the 6 degree of freedom robot via a drag-and-drop type gesture. The goal of this system is to enable a user to command a robot to perform pick-and-place actions, moving Lego blocks from one location to another. In Puljiz, Stöhr, Riesterer, **et al.** [29], a feasibility study explores a method of generating the robotic arm as a manipulable hologram within the HoloLens, using a registration algorithm and the built-in gesture recognition. The virtual robot is overlaid on the physical robot, with the goal of teleoperation. Either the end-effector can be manipulated, or the linkages can be moved to create the desired positions. In practice, issues with segmentation resulted in the hand tracking not performing well on dark backgrounds and when close to objects.

The study conducted in Elsdon and Demiris [30] uses a HoloLens in conjunction with an "actuated spray robot" for application of specific doses of topical medication. The amount of medication dispensed is shown to the user only via AR, rendering an otherwise unobservable result for the user.

Reardon, Lee, and Fink [31] show how AR can aid a human who is conducting search efforts collaboratively with a mobile ground robot. In this case the robot is providing location and navigation information to the human teammate via AR. The primary technical contribution from this study is the alignment of the frames of the human and the robot. This study also uses AR markers for testing of targets and navigation. The goal of Kästner and Lambrecht [32] is to evaluate the HoloLens's performance under 5 different visualization modes: without any sensor data visualization; with laser scan visualization; with environment map visualization; with laser scan and environment map visualization; and with laser scan, environment, and navigation visualization. The experiment uses AR to present a visual map of the space, set goal locations for the ground robot, and visualize the robot path along the floor. The main limitations of the technology are from constant visualization of real-time data, especially the laser scan data for position and obstacle tracking.

Hedayati, Walker, and Szafir [33] explore three different design methodologies, which all prove to be improvements over the baseline. A HoloLens is again utilized as the ARHMD platform, with three classifications for interface designs: augmenting the environment (which they call the *Frustrum* design), augmenting the robot (the *Callout* design), or augmenting the user interface (the *Peripherals* design). These design frameworks work quite well for the situations where the robot is separate from the human and they are collocated in the environment, but may not apply as well in all situations, for example when the robot is a wheelchair that the user is operating from a first-person perspective. In related work, Walker, Hedayati, Lee, **et al.** [18] also utilizes this design framework (augmenting the environment, augmenting the robot, augmenting the user interface), and showcases four reference designs (NavPoints, Arrow, Gaze, Utilities) for designing AR for HRC.

Limitations and drawbacks of head-mounted displays are made clear in Qian, Deguet, Wang, **et al.** [34], where a HoloLens is used to assist the first assistant during robotic-assisted surgery. The weight of the device as well as its limited field of view are both stated as problematic in participant interviews. The intent of AR in this case was to be able to (virtually) view instruments inside the patient and to provide real-time stereo endoscopic video in a convenient location.

Similarly to Qian, Deguet, Wang, **et al.** [34], Walker, Hedayati, and Szafir [20] also uses a HoloLens to display a hologram robot ("virtual surrogate") that is manipulated for teleoperation.

However, in this study the user is collocated with the robot, which is an aerial quadcopter robot instead of a tabletop robotic arm, and a handheld Xbox controller instead of hand gesture recognition is the mode of teleoperation. Two designs are tested: one which behaves like a typically teleoperated robot with the physical quadcopter immediately responding to the virtual surrogate's movements, and another where the virtual surrogate is used to set waypoints in AR which the physical quadcopter can be signaled to begin at any time. These are compared against a purely teleoperated robot, without any virtual surrogate. In the user study, both task completion time and response time are faster in the experimental conditions, and participants also preferred the experimental designs over direct teleoperation.

2.2.2 Mobile Devices: Handheld Display

Augmented reality that uses a handheld mobile device display, such as a tablet or smartphone, is a frequent implementation of AR. These kinds of devices are ubiquitous, and creating an app that can be deployed to almost anyone is relatively straightforward, simple, and inexpensive. Since the release of the iPhone in 2007, mobile devices like it are increasingly at people's fingertips, and there is already a dependable baseline level of familiarity with how to interact with AR in this form. As mentioned in the introductory paragraph to this section, handheld mobile displays provide for an AR experience that is non-immersive as compared to the HMD; furthermore, handheld devices are typically more affordable ways to implement AR for HRC.

The AR format in Fung, Hashimoto, Inami, **et al.** [35] uses the Sony Vaio ultra mobile PC, a handheld touchscreen device that recognizes fiducial markers (special tags) in the space to provide on-screen information to the user, enabling them to program a robot to carry out a limited set of tasks. The user takes photographs with the handheld device, enabling recognition of objects and locations in the photograph, and then actions are allowed to be programmed using these recognized objects and locations. In this way a robot can be programmed to operate simple home appliances, such as a hot water kettle.

The Samsung Galaxy S II smartphone is used in Lambrecht and Krüger [36], as the mobile device on which to display AR, with the goal being intuitive industrial robot programming. The mobile device displays virtual objects relevant to the robot's motions, and the user can interact using hand gestures. Information from both an external 3D motion tracking system and the 2D camera on the mobile device are combined to interpret the hand gestures.

That same year Bonardi, Blatter, Fink, **et al.** [37] present an iPad application for arranging robotic movable furniture either in situ with AR ("Augmented/A") or in virtual reality ("Virtual/V"). Tables and chairs can be placed virtually into the actual environment, and different experimental conditions either allowed the participant to move freely about the space with the iPad ("Dynamic/D") or required them to remain stationary with the iPad anchored in place ("Static/S"). Participants were also tracked with the Kinect sensor. All subjects in this 2x2 study were provided time to practice using the software on the iPad using the virtual, static condition, and then performed two of the four conditions (SV, SA, DV, or DA). Participants preferred dynamic over static conditions and performed better in the dynamic condition with respect to precision, and also expressed a preference for augmented representation over virtual despite no observed performance differences. The choice of an external mobile display for the interaction is notable here, as it allows the person to manipulate objects on a tangible screen while moving around the environment with their field of view unencumbered.

A Samsung Galaxy Tab 4 is used to compare the use of AR with traditional robot programming in an industrial environment in Stadler, Kain, Giuliani, **et al.** [38]. The participant completes three different tasks to program a Sphero 2.0 robot ball in either an AR or no-AR condition. In the AR condition, "task-based support parameters" are provided, whereas these parameters are not given in the no-AR condition. Workload measures are lower in the AR condition, while task completion time increases, possibly due to the apparent desire for participants to be more accurate in the AR condition, provided with more visibility to the task.

More industrial robot programming is explored with mobile screen AR in Hügle, Lambrecht, and Krüger [39]. The user first moves around the space with a tablet, using pointing and arm movements, while the 6-DOF robot arm remains stationary. Next the user validates robot poses and trajectories aided by the AR application, able to adjust the program as well as physically move the robot. Finally the user leaves the area so that the robot can safely demonstrate its learned movements. Gestures are recognized using the tablet's camera, the user receives AR feedback on the gesture interpretation, and a virtual robot is also displayed to demonstrate the current program.

The Apple iPad Pro is the mobile device of choice for Frank, Moorhead, and Kapila [40]. Fiducial markers are arranged on a table surrounding a humanoid robot with two 6-DOF arms. Manipulable objects, also labeled with markers, must be moved around the table. Three different interfaces, all using the iPad, are tested in a between subjects study. The three interfaces are a Conventional Egocentric (to the robot) Interface, where users view the area from the perspective of the robot's on-board camera; a Conventional Exocentric Interface, which displays an overhead camera view of the workspace; and an experimental Mobile Mixed-Reality Interface, which uses the tablet's rear-facing camera as the point of view. The reachable space can be highlighted virtually on the tablet. Statistically, participants perform equally well with all interface modes. Because the Egocentric Interface requires users to move around to gain perspective of the robot, this modality is less preferred by participants than the other two modalities. Likewise, the Egocentric Interface users also report higher workload. There is obvious variability among participants using the mobile interface, possibly due to the variety of movements available to those users.

In Sprute, Tönnies, and König [41], a Google Tango tablet with an RGB-D camera is used to define spaces that a mobile robot is allowed to occupy, using "virtual borders". Holding the tablet, a user moves around the space and chooses points in a specified plane. These points are displayed on the screen along with the virtual borders which they define. This method is compared against two baseline methods: visual (physical) markers and a laser pointer. Ultimately the results showed that the tablet method produced similar accuracy as the baseline methods and resulted in a faster teaching time.

In Chacko and Kapila [109], a Google Pixel XL allows a user to select an object and a goal location, which are then shared with a 4-DOF tabletop robot manipulator with a 1-DOF gripper. The mobile AR display features two buttons (one for setting the target and another for clearing), crosshairs to assist with locating a target, shading to denote reachable regions, and virtual objects to indicate intended final placement. Different versions of the interface are provided to allow the user to program either one pick-and-place object at a time or multiple objects together. Participants rate the workload required for this task and interface as relatively low. Chacko and Kapila [42] extend Chacko and Kapila [109] by expanding the types of objects to be manipulated, allowing for two different grasping modes (vertical and horizontal), and adjusting the AR display accordingly.

The software developed in Rotsidis, Theodorou, Bryson, **et al.** [43] is intended to facilitate trust between robots and users, using a mobile phone AR application to increase transparency. The AR display has modes that show a ground robot's decision-making capabilities in tree-like formats. Subtrees can be expanded with a tap, and users can debug the program and access additional information. This kind of transparency increases the likelihood that the robot is perceived as alive, lively, and friendly by study participants.

As demonstrated by this chronological review of mobile device AR display, the uses are incredibly diverse and allow for a variety of functionality and information provision.

2.2.3 Projection-based Display

Another commonly used mode of augmenting the real world for HRC is projection. Much of the work in this area has occurred within the past 4 or 5 years, perhaps due to the maturation of projection and motion capture technologies.

In 2016, work in Andersen, Bøgh, Moeslund, **et al.** [44] utilizes projection mapping to facilitate autonomous robotic welding. An operator uses a Wii remote to control a cursor and communicate with the robot. In the experiment, the projection is displayed on a mock-up of a shop wall. The participant completes two separate tasks, one requiring them to correct a number of incorrect locations for welding, and another to teach the welding task to the robot. The functionality of the projection system was rated relatively highly by mostly novice participants, due in part to the projection visualization of task information.

In a car door assembly task Kalpagam Ganesan, Rathore, Ross, **et al.** [45], projections are used to dynamically indicate various cues to human collaborators with robots. Object locations are tracked with a vision-based system, and this enables projection mapping on top of the 3D objects. Three modes of communication were tested: printed mode, in which subjects received printed instructions; mobile display mode, in which subjects received a tablet with instructions; and projection mode, providing just-in-time instructions via projection mapping with mixed reality cues. Participants had to collaborate with a robot to complete the door assembly task. The amount of time required to understand a subtask was lower in the projection mode than in the printed or mobile display modes. Furthermore, the subjective questionnaire revealed higher fluency, clarity, and feedback with the projection mode. All participants also favored the projection mode in this within subjects test.

In another industrial application in Materna, Kapinus, Beran, **et al.** [46], a human subject uses spatial augmented reality to program a robot to prepare parts for assembly. Projections are displayed on a touch-enabled table that is also within reach of the robotic arms. Since all work occurs on the table, the location of the projections in this same area is intended to increase focus and situational awareness, improve use by novice users, and remove the need for other devices. The tabletop system serves both as input for the robot and feedback for the human. Lists of instructions and programs, dialog boxes, and images representing objects to be manipulated are all "widgets" shown on the tabletop surface. Unfortunately, the affordances of the touch-capable table proved to be lacking, and 5 of the 6 participants agreed with the statement, "Sometimes I did not know what to do," demonstrating once again that shortcomings in the tools can deeply affect the overall experience.

Similar to Materna, Kapinus, Beran, **et al.** [46], in Bolano, Juelg, Roennau, **et al.** [47] a tabletop projection system is also used. In this case, however, information is shown about robot behavior and detected parts, with the goal of clarifying the task and the robot's intent, and the table is not touch-enabled, nor are any inputs solicited from the user. Without the hindrance of a confusing touch interface as in Materna, Kapinus, Beran, **et al.** [46], the usefulness of tabletop projection can be assessed. Because in this example the user is working concurrently with the robot rather than programming it, understanding intent and future movements is especially useful. If the robot makes an unpredictable move, the human user can see with a glance the goal location and immediately assess whether or not a collision is imminent.

2.2.4 Static Screen-based Display

A mode of AR display that has declined in popularity in recent years is that of a screen-based display, generally placed on a desktop for viewing. This display is distinct from the mobile device displays discussed earlier, as it cannot be moved with the user on the fly, nor is it generally equipped with a mobile camera. Research involving static displays for HRC is largely for remote use purposes, featuring an exocentric camera view and virtual overlays for the remote user. Here we highlight some examples of these static displays for AR, though this modality has been less common in recent years.

Work in 2009 used a screen-based display to facilitate dental drilling in Ito, Niwa, and Slocum [48]. Virtual images were projected onto teeth to perform the drilling required to prepare them for a crown. The path of the drill can be superimposed, and feedback shown on the screen. The machine is teleoperated via joystick, and the AR system enables replication of the original operation.

In 2010, a remote operator is shown a live view of a robot arm with additional information on top of and around the robot in view in Notheis, Milighetti, Hein, **et al.** [49]. Both virtual and real cameras are enabled, with the virtual model showing the intended movement of the real robot. The user can validate the movements via the screen prior to the action being taken in real life.

In proof-of-concept work done in 2012 in Domingues, Essabbah, Cheaib, **et al.** [50], the intent is to provide users with a virtual scuba diving experience. While an underwater robot (ROV) was teleoperated, a screen-based AR displays controls and the video feed from the ROV. The user can choose whether to use the on-board ROV camera or the virtual ROV for controlling the robot.

A stationary touchscreen AR display is used in 2013 to allow users to teleoperate a ground-based robot in another room by manipulating a 3D model on the screen in Hashimoto, Ishida, Inami, **et al.** [51]. The user draws the robot path on the screen with their finger, and various cameras are provided to augment the user's view, including a third-person view camera. Three movement modes are tested with the touchscreen input: Movement After Touching (the robot does not move until the person is no longer touching the screen), Movement During Touching (the robot moves as soon as the user begins to manipulate the model but stops immediately when the screen is no longer being touched and the model moves to the current location of the robot), and movement during and after touching (the robot begins as in Movement During Touching, however when the user stops touching the screen, the robot continues to the final model position). Only 12 participants were involved in the study, which makes generalizations about the usefulness of each mode difficult, and there were participants who preferred each of the three modes.

2.2.5 Alternate Interfaces

A survey of literature in AR for HRC would be deficient without the acknowledgement of the development of various peripheral devices for interacting in augmented reality. Here we provide examples of the diverse types of peripherals.

One example of a peripheral being used with AR is in Osaki, Kaneko, and Miwa [52], where a projection-based AR is combined with a drawing tool peripheral to set a path for a mobile ground-based robot. Additional commands and communication are provided by the drawing tool including navigation by virtual string (as if it were a leash and the robot were a dog) and the use of different colors to indicate stop or go.

To enable robot use by people with mobile disabilities, a "tongue drive system" (TDS) is developed for use with an AR headset in Chu, Xu, Zhang, **et al.** [53]. Using tags and object recognition, a user is able to perform pick-and-place and manipulation tasks faster with the TDS than with manual Cartesian inputs from a keyboard.

One proposed concept, and an example of where this kind of technology might lead us in the future, is an immersive suit for the elderly: the "StillSuit" in Oota, Murai, and Mochimaru [54]. The main purpose of the robotic StillSuit is to enable interaction with the environment. Using "Lucid Virtual/Augmented Reality," the central nervous system and musculoskeletal system are modeled, providing the user with the sensations of performing a particular task.

In Gregory, Reardon, Lee, **et al.** [55], users perform gestures while wearing a Manus VR gesture glove, capable of tracking each finger's movement. While wearing a HoloLens, users provide movement instructions to a ground-based robot via the gesture glove. A key insight learned in this pilot study is that gestures should be chosen so that they can be easily formed by all users.

2.2.6 AR Combinations and Comparisons

Other themes in the literature included the comparison of different AR modalities via user studies and the combining of modalities to achieve improved effects. These studies bear importance for those who may be deciding whether to implement AR in different modalities or how to provide AR insight to both an egocentric and an exocentric user simultaneously, thus related works are shared below.

Augmented reality can be a combination of technologies, such as in Huy, Vietcheslav, and Seet Gim Lee [56], which combines projections using a laser writer system (or *spatial augmented reality*, SAR) with the Epson Moverio BT-200 AR Glass (an HMD) and a multimodal handheld device prototyped for the study. The laser writer is mounted to a ground-based mobile robot to provide directional feedback, the human can provide commands via the handheld device, and other visual feedback can be provided via the HMD. The intent of testing both versions of AR (projection and HMD) is for those cases where some of the communicated information may be sensitive, while other information may be needed by all those in the vicinity of the robot for safety purposes.

Sibirtseva, Kontogiorgos, Nykvist, **et al.** [57] compare different AR methods where the three conditions are HMD, projector, and a monitor. Participants claim that the HoloLens is more engaging, possibly due to the mobility that an HMD allows, but generally prefer the projection-based AR for a tabletop robot manipulator conducting a pick-and-place task because it was "natural," "easy to understand," and "simple."

Similar to Huy, Vietcheslav, and Seet Gim Lee [56], in Bambušek, Materna, Kapinus, **et al.** [58] a HoloLens is combined with projection AR, so that an outsider can see what the HMD-wearer is doing. The study indicated a high task load for the HMD and confusion when both were used. Ultimately the task completion time was faster with the HMD regardless of the high Task Load Index rating. The unreliable touch-enabled table proved to be problematic, as seen in other studies like Materna, Kapinus, Beran, **et al.** [46].

AR (and VR in this instance) have also been used as training tools for operation of a conditionally autonomous vehicle in Sportillo, Paljic, and Ojeda [59]. In a between-subjects study, three different training methods are tested: on-board video tutorial, AR training, and VR simulator. In this wizard-of-oz study, all participants are able to take over in the appropriate situations within the required time, regardless of their training method, but participants trained with AR or VR have a better understanding of the procedure and better performance time.

2.3 Creating and Understanding the System

We encountered a large subset of literature that discussed the problems of allowing a user or designer to better understand, create, or improve the human-robot collaborative system via augmented reality. Below we discuss these in respective subsections based on the ways in which they do so or their intended domain.

2.3.1 Intent Communication

Research highlighted in this subsection addresses the problem of communication of robot intent to humans via AR. The following section, 2.3.2 Path and Motion Visualization, is related to intent, but it is differentiated in that intent is not always path- or trajectory-based. A robot might want to communicate an overall plan, a goal location, or a general intent so that the human collaborator does not duplicate efforts, alter the environment, or put themselves in danger. Thus, we share this section specifically dedicated to intent communication.

One key example of intention explanation is in Chakraborti, Sreedharan, Kulkarni, **et al.** [60], where the "Augmented Workspace" is utilized both before and during task execution. The aim of this work is to keep the human collaborator informed, increase the fluency of the collaboration, increase clarity of the plans (before and during task execution), and provide a common vocabulary. Particularly notable is the Projection-Aware Planning Algorithm, where “the robot can trade-off the ambiguity on its intentions with the cost of plans.” Similarly, algorithms for interpreting the scene and establishing and updating the virtual borders to be shown to the HMD wearer are presented in Sprute, Viertel, Tönnies, **et al.** [61].

The overarching goal of Reardon, Lee, Rogers, **et al.** [62] is to provide straightforward, bidirectional communication between human and robot teammates. The human is provided information to more clearly understand the robot's intent and perception capabilities, while the robot is provided information about the human that enables it to build a model. By enabling this bidirectional communication, the authors seek to influence human behavior and increase efficiency of task completion. The task at hand in this experiment is the cooperative exploration of an uninstrumented building. The robot and human (wearing an AR HMD) are independently performing SLAM, and their frames of reference must first be aligned with each other. Next the maps from both sources are composited. Finally the robot's information is provided to the human teammate visually, in their AR-HMD. Information visually communicated to the human via the AR-HMD includes: the robot's current plan; the composite map, to facilitate understanding of the current state of the exploration task; and other information to convey how the robot is evaluating future actions [62].

In cases where humans and industrial robots must work in close proximity, safety and trust can be improved by indicating the robot's intent to the human. For example, in Bolano, Juelg, Roennau, **et al.** [47], a human collaborator works in a shared space on an assembly task. Using projection-based AR, the user can immediately see whether a part is recognized by the system and also be shown the current target, trajectory path, and/or swept volume of the robot, so that they can safely move out of the way (or know that they are already working in a safe space), even if it might appear as though the robot is moving towards them.

To aid in the disambiguation of human commands, Sibirtseva, Kontogiorgos, Nykvist, **et al.** [57] present a system that involves natural language understanding, a vision/object recognition module, combining these two for reference disambiguation, and the provision of both a visualization in AR and an autonomous robot controller. After a pilot study to establish human language preferences for the reference disambiguation visualization system, a relatively straightforward pick-and-place task for different colors of blocks is established to compare three modalities of AR.

In a similar experiment, Williams, Bussing, Cabrol, **et al.** [63] performs a within-subjects study to investigate how a robot can communicate intent to a human via AR images as deictic gestures (such as circling an object in the user's field of view), rather than using physical deictics (such as pointing). The experimental results suggest design guidelines for "allocentric mixed reality deictic gestures," including the suggestion to use these gestures in contexts where language may be difficult or impossible, or when the intended target may be perceived as outside the robot's perspective, and to use them in combination with language when the situation allows.

A key result of communicating robot intent is the calibration of a human user's trust that results from their mental model of the system and from an understanding of its capabilities and limitations. This calibration of trust is one of the primary goals of Rotsidis, Theodorou, Bryson, **et al.** [43]. Using a mobile phone-based AR, a tree-like display of the robot's plans and priorities was shown to a human for both transparency and for debugging.

Even more recently, [64] compared different two different AR robot gestures (a virtual robot arm and a virtual arrow). Based on the robot's deictic gesture, the participant chose the virtual item that they believed the robot was indicating. While the arrow gesture elicited more efficient responses, the virtual arm elicited higher likability and social presence scores for the robot. These results carry various implications for intent communication, including an important choice between likability and efficiency. Further, AR is shown in [65] to be a a promising technology for bi-directional communication of intent and increased task efficiency through experiments that provide avenues for both the human and the robot to communicate intent and desires. Other AR-enabled indication methods that have been explored include a virtual robotic arm on a physical robot that points to desired objects, as demonstrated in Hamilton, Phung, Tran, **et al.** [64]. This study compares the virtual arm with a virtual arrow, and finds that while arrows support a faster reaction time a virtual arm makes the robot more likable. AR-based visualizations that include placing a virtual robot in the physical space along with sensor data and a map grid are also tested in Ikeda and Szafir [66] for supporting debugging by roboticists.

2.3.2 Path and Motion Visualization and Programming

Another popular problem in human-robot collaboration is that of understanding and programming robot trajectory and motion. As clarified in Section 2.3.1, here we focus on paths and trajectories of the robots, and how AR can be used to visualize or program these trajectories.

In a straightforward and intuitive example from Osaki, Kaneko, and Miwa [52] in 2008, the human user draws lines in AR (via both projector and HMD), using a peripheral device, for the robot to follow. The lines are then processed into trajectories which the robot can take. Similarly, in Chestnutt, Nishiwaki, Kuffner, **et al.** [17] a human user directs a humanoid robot by drawing a guide path on the ground in AR. The system then plans left-right footstep sequences for the robot that are also displayed via AR, and the user is able to modify the path if necessary.

For a remote laser welding task, a similar line-following approach is taken in Reinhart, Munzert, and Vogl [67], also in 2008. First the welding locations are denoted with the specific welding task to be completed using AR projections, and next the robot paths are optimized for task completion. Approximately 8 years later, Andersen, Bøgh, Moeslund, **et al.** [44] is also related to welding, this time for stud welding in a shipbuilding environment. Projection mapping is used in this instance as well, and a lab-based user study indicates positive results for novice users in programming the robot to conduct accurate welding activities.

In Green, Chase, Chen, **et al.** [21], the authors set three different experimental conditions for humans navigating a simulated robot through a maze with the use of AR. The 3 within-subjects conditions tested are: Immersive Test, using an onboard camera and teleoperation without any AR; Speech and Gesture no Planning (SGnoP), providing AR interaction with speech and gesture; and Speech and Gesture with Planning, Review, and Modification (SGwPRM), adding to the prior condition the opportunity to review the plan before it is executed by the robot. While the Immersive condition is preferred by test subjects and most easily executed, SGwPRM yields the most accurate results. Significant user learning had to take place in both of the AR conditions, while the pure teleoperation is a more natural mode of control. This study combines a number of different options, such as displaying the path before robot movement begins, utilizing AR tags to display virtual objects to the user, and integrating speech and gesture inputs.

A significant amount of research covers different ways to “teach” or program a robot using AR. Here we present them chronologically, in part to highlight the progression of the research over time.

In Hulin, Schmirgel, Yechiam, **et al.** [68], visual and haptic signals are given to a human via AR who is using Programming by Demonstration to teach a robot arm a trajectory. The signals are intended “to avoid singularities”. The following year in Fung, Hashimoto, Inami, **et al.** [35], a human user takes photographs with an AR-enabled device and then provides annotations, which transfer to a ground robot’s movement. In another study from Bonardi, Blatter, Fink, **et al.** [37], while it does not use separate ground robots, the furniture itself is robotic and modular. Users interact with an iPad to control the arrangement of the furniture in a shared space. While these papers covered scenarios with humans in the same space as a robot, Hashimoto, Ishida, Inami, **et al.** [51] instead deals with a robot being teleoperated from another room via touchscreen. Also in 2013, Gianni, Gonnelli, Sinha, **et al.** [69] present a framework for remotely operating a semi-autonomous ground robot as well. Their framework includes an AR interface that allows for path planning and obstacle navigation through a handheld pen peripheral, as well as a localization system that used dead reckoning in addition to ICP-SLAM, and a trajectory tracking algorithm. This kind of remote communication is designed to be especially useful for situations that might pose greater risk to a human, such as emergency rescue or scouting. Both Lambrecht and Krüger [36] and Lambrecht, Walzel, and Krüger [70] focus on honing hand gesture recognition algorithms for spatial programming of industrial robots. Specific contributions include recognition of specific gestures that map to robot poses, trajectories, or task representations, and improvements in the skin color classifier and hand/finger tracking. In a 2014 user study, Coovert, Lee, Shindev, **et al.** [71] demonstrate the effectiveness of projections (such as arrows) from the robot onto the floor in front of it when moving in an environment among humans. Participants feel more confident about the robot's movement and more accurately predict its movement with projections than without. In another study the following year, Chadalavada, Andreasson, Krug, **et al.** [72] suggest that a mobile ground robot that projects its intentions onto the floor with simply a contour line is preferable to no projection at all.

Rather than use AR for directing or programming the robot, Makris, Karagiannis, Koukas, **et al.** [73] suggest that an AR HMD can be used in a human-robot collaborative assembly environment to provide the human with robot trajectory visualizations, so that they can stay safely away from those areas. However, the presented system does not offer any recourse if the user does intersect the denoted trajectory/path. In a study by Walker, Hedayati, Lee, **et al.** [18], different ARHMD visualization designs are tested for communicating to a human in a shared space what the intent of a quadcopter robot is. Four different visualizations are tested in a between subjects study: NavPoints, Arrow, Gaze, and Utilities. These visualization designs each have different purposes and uses.

Hügle, Lambrecht, and Krüger [39] present a programming method for a robot arm that involves both haptic (Programming by Demonstration) and gesture-based input. The gesture-based input is used to provide a rough definition of the poses within the space, while AR images are used to validate the poses and trajectories and alter the program. Next, the human takes turns leaving the space while the robot moves to the next pose, re-entering the space to provide hands-on feedback and alterations, and then leaving again for the next movement. Once the program is finalized, it is transferred to the controller.

In Materna, Kapinus, Beran, **et al.** [46], users program a PR2 robot as an assembly assistant, using projection-based AR on a touch-enabled table. They use a block programming technique (with the blocks projected on the table) to select the appropriate steps for the robot to complete, and the target locations for parts are also highlighted virtually on the table. Templates are available offline for the users to work from, and specific parametric instructions (such as *pick from feeder* or *place to pose*) are supported. No pre-computed joint configurations or trajectories are stored, and all paths are planned after the program is set.

The system in Krupke, Steinicke, Lubos, **et al.** [74] allows a human user to interact with a virtual robot, move it virtually, confirm the movements via speech after watching a visualization of the picking motion, and then observe the actual physical robot move according to those movements, the goal being a pick-and-place task. In another pick-and-place task, non-experts are asked to program a robot used to move printed circuit boards to and from their testing locations [75]. A form of block programming is used in which "pucks" are chosen and placed by the user to indicate actions and their sequences to the robot. Bambušek, Materna, Kapinus, **et al.** [58] provide a user with a HoloLens HMD for programming a robot for a pick-and-place task, but also augment it with AR projections so that others can see what the HMD-wearer is doing, to avoid confusion and provide for safety. In this case, the robot need not be present for the programming to take place, as object placement occurs entirely virtually at first. Interactive Spatial Augmented Reality (ISAR) occurs along with *virtual* kinesthetic teaching (ISAR-HMD).

In Kästner and Lambrecht [32], a large portion of the work focuses on aligning the coordinate systems of the HoloLens and the robot, similar to Reardon, Lee, Rogers, **et al.** [62], both in 2019. After alignment is assured, then sensor data can be visualized, which includes the navigation path of the robot that is extracted from the global path planner. Results show a struggle to visualize the large amounts of real-time laser scan data using the HoloLens, a limitation to be addressed in the future. To assist humans in remotely exploring unsafe or inaccessible spaces via UAV, Liu and Shen [76] use a HoloLens to display an autonomous UAV's "perceived 3D environment" to the human collaborator, while the human can also place spatial targets for the robot. In an attempt to develop an all-inclusive AR system, Corotan and Irgen-Gioro [77] present a combined augmented reality platform for "routing, localization, and object detection" to be used in autonomous indoor navigation of a ground robot. Other noteworthy recent research presents AR-based methods for programming waypoints and states for robot arms [78], [79], as well as for programming robots through learning from demonstration [80], and for projecting intended paths a social robot might take [81].

2.3.3 Adding Markers to the Environment to Accommodate AR

One method of making AR easier to implement is to change the surroundings by providing tags, markers, or other additions and alterations. While this requires that the environment can actually be prepared in this way (both that it is physically possible and temporally feasible), these kinds of features can significantly increase the ease of AR implementation. Furthermore, AR markers and tags are generally used to address problems of placement, labeling, and recognition encountered when using AR technology, and aim to increase user understanding of the system. Below we share research that demonstrates these kinds of accommodations, again chronologically.

In Green, Chen, Billinghurst, **et al.** [82], a Lego Mindstorms NXT robot path is planned by a human user by combining fiducial markers, other graphics, gestures, and natural language, specifically deictics. Paddles with different markers that indicate instructions such as "stop" or "left" provide instructions for the robot, while the robot confirms the human's plan using natural language responses. AR, specifically using the markers in the environment, allows for a common communication platform between the human and robot. The exploration of AR for HRC using AR markers continues to progress in Green, Chase, Chen, **et al.** [21], where the authors set three different experimental conditions for humans navigating a simulated robot through a maze with the use of AR. AR markers are placed in the participant's physical environment, on which the virtual obstacles in the maze were modeled.

A similar task of programming a robot to follow a pre-set list of instructions utilizes fiducial markers in Fung, Hashimoto, Inami, **et al.** [35]. With this handheld AR, labels are displayed in the user's view, allowing them to match the objects with the provided instructions, and then provide direction to the robot.

The title of "Mixed reality for robotics" in Hönig, Milanes, Scaria, **et al.** [83] is so generic as to give away the novelty of this research area. The authors' goal is to show how mixed reality could be used both for simulation and for implementation. One single physical robot is used as a basis for additional virtual robots, and simulation is pitched as a research and development tool. In this study, markers are placed on the robots in the real world to make it easier for the simulation to mimic the motion directly.

AR has been explored for many uses in a manufacturing environment, such as in Peake, Blech, and Schembri [84] where AR markers are used to overlay objects on the factory floor. The images displayed virtually can be pulled from the cloud and can provide information about machine status and equipment usage.

There are many kinds of uses for AR tags and fiducial markers, or ways in which the environment can be altered to accommodate the use of augmented reality. Fiducial markers are used in Frank, Moorhead, and Kapila [40] to both denote possible goal locations and to label movable objects, which are to be recognized by the robot and the AR device. This simplifies the recognition aspects significantly, removing that process from the system. In order to locate and orient a ground-based robot in a confined space, Hashimoto, Ishida, Inami, **et al.** [51] label its corners with fiducial markers. This facilitates the control of the robot by a remote user via touchscreen.

2.3.4 Manufacturing and Assembly

One domain in which solutions for creating and understanding the human-robot collaborative system are particularly applicable is that of manufacturing and assembly. Specific tasks performed in such environments, and which can benefit from the use of AR, include tool alignment, workspace visualization, safety precautions, procedure display, and task-level programming. Especially over the last 5 years, the manufacturing environment has become a popular research area for AR in HRC.

In a study intended to represent the tasks of a factory robot, Stadler, Kain, Giuliani, **et al.** [38] task participants with using a tablet-based AR to teleoperate a Sphero robot in 3 different activities: tool center point teaching, trajectory teaching, and overlap teaching. The AR tablet provides "task-based support parameters" in the form of shapes, guiding lines, start and end points, and radii. Workload decreases with the tablet-based AR, however task completion time increases. The authors suggest this could be attributed to the support parameters providing a visible comparison for exactness.

In a robot-assisted assembly scenario, AR shows potential usefulness in multiple ways, such as displaying assembly process information, visualizing robot motion and the workspace, providing real-time alerts, and showing production data [73]. The specific case study applies to the automotive industry, where a COMAU NJ 130 robot works in a cell collocated with a human. A red volume denotes the robot's workspace, the green volume is safe for the operator, and the current task is shown at the top of a screen. This proof of concept is intended to show the additional safety and efficiency afforded with the use of AR. Also in 2016, [85] apply an "object-aware projection technique" to facilitate robot-assisted manufacturing tasks like the installation of a car door. Projections such as wireframes and warning symbols aid the human in understanding robot intent. Another study intended to improve assembly operations, Materna, Kapinus, Beran, **et al.** [46] uses a PR2 robot as the worker's assistant, helping to prepare the parts for assembly. The worker is aided by AR to create a block program for the robot, see the instructions, view object outlines, and receive information about the state of the system as well as additional information. Unfortunately the robot itself is relatively unreliable during the experiment, and other usability issues are also apparent (participants blocking part of the table where the robot should place its parts, or participants intentionally or unintentionally ignoring errors shown via dialog boxes and audio in the system). Future studies should take into consideration these kinds of limitations.

[84] also work towards implementing AR in a robot-enabled factory, using a mobile device and AR tags to display virtual objects and their expected manipulation by the robot on the factory floor. Research in Guhl, Tung, and Kruger [24] takes this concept further by implementing multiple AR modalities that allow a worker to impose movement restrictions, change joint angles, and create programs for a robot in the factory on the fly, including the UR 5, Comau NJ 130, and KR 6.

A seemingly common application for AR for HRC is in robotic welding [25], [44], [67]. The dangers of welding combined with the accuracy required for welding tasks are perhaps what make this a potentially useful application. In Reinhart, Munzert, and Vogl [67], AR was used to assist with programming the remote laser welder, providing a user the capability to define task-level operations. In both Reinhart, Munzert, and Vogl [67] and Andersen, Bøgh, Moeslund, **et al.** [44], projection-based AR is used to display the weld plan to the user directly on the area to be welded. In Yew, Ong, and Nee [25], however, an HMD displays virtual objects in the user's field of view so that they can teleoperate a remote welder.

Puljiz, Krebs, Bösing, **et al.** [86] draw on the built-in mapping and localization capabilities of the HoloLens to establish safe zones and other areas of interest within a robot cell, rather than relying on an external source. Results presented in the paper show that the mapping can aid in setup of the robot cell, and the HMD allows for straightforward editing of the map and safety zones.

2.4 Improving the Collaboration

The subsections that follow contain literature that addresses the problem of improving the collaboration between the robot and the human via augmented reality. Research is grouped depending on the domain of the collaboration. We examine domains from different perspectives, including use cases and applications.

2.4.1 AR for Teleoperation

Beginning with [110] and continuing with [111], robot teleoperation has remained a central problem in human-robot collaboration, for which augmented reality can provide some solutions. The contributions of research using AR for teleoperation are chronologically summarized here.

Ito, Niwa, and Slocum [48] suggest visual overlays for robot-assisted, teleoperated dental work, in yet another example of the use of AR for HRC in the medical fields. In this particular case, the work is not done directly on patients but for a dental milling machine to prepare tooth crowns. In this paper, the machine itself is presented, with the AR concept being a virtual object superimposed over the actual object while the machine was being operated.

For UAV (unmanned aerial vehicle) control, AR has been shown to improve the situational awareness of the operators and to improve the path choice of the operators during training as in Hing, Sevcik, and Oh [87]. (For more on situational awareness evaluation, see Section 2.5.1.5.)

Operators are provided with two different types of AR "chase views" that enable them to observe the UAV in the environment. Other teleoperated robots are those operated beneath the surface of the water (ROVs, or remotely operated vehicles, also known as UUVs or unmanned underwater vehicles). Domingues, Essabbah, Cheaib, **et al.** [50] present a virtual diving experience that used teleoperated ROVs and AR. Riordan, Horgan, and Toal [88] showcase a real-time mapping and display of subsea environments using technology enabled by UUVs; this provides remote teleoperators with a live experience of the environment in relatively high resolution via the combination of technologies presented in the paper.

Another way of assisting a remote operator is by placing them virtually into the environment of the robot as in Krückel, Nolden, Ferrein, **et al.** [23], so that they can in fact operate egocentrically. An alternative to placing the operator into the entire virtual environment is to use a combination of virtual and real objects to mimic the robot's workspace, as in Yew, Ong, and Nee [25]. In this example, a maintenance robot is shown virtually in AR, along with some aspects of its surroundings, while prototypes of some of the physical features are also present in the operator's immediate environment. In this way, tasks such as visual inspection or corrective task execution can be completed remotely via teleoperation.

With the comprehensive system presented in Huy, Vietcheslav, and Seet Gim Lee [56], a peripheral/haptic device is used to teleoperate the robot, and information and feedback are shown to the human user via an HMD and laser projection mounted to the mobile ground robot. One feature of the handheld peripheral is a laser pointer that can be used to identify a goal location for the robot, following which the operator confirms the choice in AR, then the robot moves to that location autonomously.

As the concept of using AR for teleoperation continues to evolve, the designs have become more advanced. In Hedayati, Walker, and Szafir [33], three different design methodologies are presented for communicating information to an operator collocated with an aerial robot. This design framework urges the designer to consider how information is presented, whether it is (1) augmenting the environment, (2) augmenting the robot, or (3) augmenting the user interface. In the experiment, each of these three interface design implementations prove to be an improvement over the baseline.

Puljiz, Stöhr, Riesterer, **et al.** [29] present a method of generating a 6-DOF robot virtually in AR with a HoloLens, and then allowing the user to manipulate the hologram as a form of teleoperation, either in situ or remotely. Similarly, Walker, Hedayati, and Szafir [20] successfully demonstrate the use of “augmented reality virtual surrogates” of aerial robots that can be manipulated using an HMD as a form of teleoperation. In a shared control situation, where a human user with a remote control must grasp an object with a robot arm using an assistive controller, Brooks and Szafir [89] show that AR visualization increases acceptance of assistance as well as improves the predictability rating, but does not affect the perceived usability. There is even evidence that humans in remote control of robot swarms prefer trajectory information delivered via AR [90].

2.4.2 Pick-and-Place

While pick-and-place operations are applicable across many of the domains already discussed such as path planning, manufacturing, and teleoperation, here we highlight problems of pick and place in human-robot collaboration as solved by augmented reality for those who are interested in this particular body of research.

In Hashimoto, Ishida, Inami, **et al.** [51], a multi-DOF robot arm is mounted to a mobile ground robot, giving the resulting system a total of 6 DOF. This robot is then teleoperated through a touchscreen AR interface to perform tasks remotely (in another room), such as approaching a bottle, grasping it, and dropping it into the trash. The experiment is designed to determine subjects’ preferred type of interaction with the touchscreen. Unfortunately these results are somewhat inconclusive, as the study was conducted on a small scale and participants did not show one clear preference.

In Frank, Moorhead, and Kapila [40] a tabletop two-armed robot is controlled via an AR-enabled tablet in a shared space. Different views are provided to the user in a between-subjects study: overhead, robot egocentric, and mobile (using the rear-facing camera on the tablet). Mixed reality is enabled in all of these views, to the extent possible with the cameras employed. The pick-and-place task requires users to command the robot to move tabletop objects from one location on the table to their designated bins on the table in front of the robot. Yet again the results show a relatively equal performance level among participants, regardless of the view provided.

Sibirtseva, Kontogiorgos, Nykvist, **et al.** [57] use verbal commands for a YuMi robot performing object retrieval tasks, and investigate the implementation of different visualizations to clarify the requests. In a within-subjects study, three visualization modalities are tested: monitor, which uses an external screen to highlight the potential object; projector, wherein the object is highlighted directly on the workspace; and head-mounted display, where a HoloLens highlights the object virtually in the real world. The system uses a wizard to perform the natural language recognition for colors and shapes of the objects; the remainder of the system is designed for the experiment. The authors choose a flat workspace for the experiment, assuming that a more complex workspace or area would essentially bias the results towards an HMD being preferable, due to difficulties with projection and/or occlusions. The claim is that this experiment is intended to compare the three AR modalities as directly as possible, rather than optimize for a specific task. While participants claim that the head-mounted display is more engaging, they generally prefer the projection-based AR.

To investigate the use of "drag-and-drop" in AR to program a UR5 robot arm, Rudorfer, Guhl, Hoffmann, **et al.** [28] test their "Holo Pick-n-Place" method. A user can virtually manipulate an object from one place to another within the HoloLens, and those instructions are then interpreted by the system and sent to the robot. The HoloLens uses object recognition to overlay the virtual CAD models of objects onto the physical objects, which the user can then drag and drop into the desired locations. A proof of concept is presented, and accuracy proves to be limited due to the HoloLens's limitations in gaze and calibration. The system also does not allow object stacking or placement anywhere other than on one surface. With the release of the HoloLens 2, some of these issues may be resolved in future studies.

In Chacko and Kapila [91], virtual objects are created and manipulated by a human user in AR, and these virtual objects are then used by the robot to optimize a pick and place task.

The system allows an estimation of position, orientation, and dimension of an object in physical space that is unknown to the robot, and this information is used by the robot to then manipulate the object. The user also dictates what type of grasping motion to use, with the options being horizontal (objects that can be grasped from above, so as to keep them oriented horizontally) and vertical (objects that can be grasped from the sides, so as to keep them oriented vertically).

In Bambušek, Materna, Kapinus, **et al.** [58], a HoloLens and touch-enabled table with AR projection are combined to program a robot to perform tabletop pick-and-place tasks. In this case, these modalities were compared with kinesthetic teaching, or physically manipulating the robot's arms. An advantage of this system is the removal of the requirement that the robot be present during programming, since tasks can be verified in the HoloLens.

2.4.3 Search and Rescue

Search and rescue operations present a natural application for using AR to facilitate and amplify human-robot collaboration. Dangerous situations can be explored by robots while a human provides guidance, oversight, and even teleoperation from a distance, using the improved situational awareness and nuanced communication enabled by AR. Specific issues that can be addressed by AR in a search and rescue HRC situation include a potentially dynamic and unknown environment, often resulting in the need for visual assistance, as well as remote communication of essential information about safety, terrain, or location of human and robot agents.

In 2009, Martins and Ventura [92] implement a rectification algorithm for using an HMD to teleoperate a mobile ground robot. In this application, head movements can be tracked and utilized to tilt the camera or turn the robot. Additionally, when the user's head is tilted from side to side, the rectification algorithm ensures that the remote image stays aligned with the horizon. Gianni, Gonnelli, Sinha, **et al.** [69] propose a framework for planning and control of ground robots in rescue environments. A human operator uses an AR interface that provides capabilities for path planning, obstacle avoidance, and a pen-style interaction modality. The following year, in 2014, Zalud, Kocmanova, Burian, **et al.** [93] demonstrate a method of combining color and thermal images in AR especially for use cases with low visibility as in rescue situations. Four years later, Reardon, Lee, and Fink [31] implemented AR for search and rescue with a ground based robot (Clearpath Robotics Jackal) using a HoloLens. The advances with this new technology included vector-style visualization of the robot pose and trajectory and expedited communication of search results.

In Reardon, Lee, Rogers, **et al.** [62], an explorer robot and human user communicate with each other via an AR HMD, with the key components being an unstructured, uninstrumented environment and bi-directional communication. An autonomous robot searches the environment with a human, with the intent to expedite the search over what could be done with solely robotic or solely human exploration. The human (via the HMD) and the robot are equipped with SLAM capability and are able to share their respective information with each other, and thus create a composite map of the area. Furthermore, the AR is used to communicate the current plan, the task's state, and future actions of the robot, thereby also influencing the choices that the human makes. In an extension of this work, Gregory, Reardon, Lee, **et al.** [55] demonstrate the usefulness of a gesture glove for giving commands to the robot for reconnaissance style missions. In a pilot study, novice participants must use the Manus VR gesture glove and a HoloLens to command the robot in mapping three different environments (subway platform, basement, and office building). Preliminary results show that these tasks can be completed both in Line-of-Sight and Non-Line-of-Sight operations without extensive training, and also highlighted the importance of choosing easily articulated gestures. Researchers also note that the participants make use of commands in unanticipated ways, such as utilizing a "return" command to only partially move the robot back, to then be able to issue a different command from this intermediate location. Reardon, Haring, Gregory, **et al.** [94] demonstrated that an ARHMD could be a suitable method for communicating robot-observed changes in the environment. The experiment, conducted remotely, provided participants with video of the environment with AR-provided, circular shaded regions that highlighted changed areas. Participants were then asked to rate their confidence in the AR-provided change indicators. While improvements could be made on this method, it proved to be a significant step in implementing this kind of visualization to aid in scene change identification. Taking these techniques a step further, Walker, Chen, Whitlock, **et al.** [95] show that an ARHMD could be used to allow emergency responders to quickly visualize an area, for example during firefighting operations, particularly by augmenting images provided by a remote robot.

Even more recently, Tabrez, Luebbers, and Hayes [96] explored different types of AR communication for joint human-robot search tasks, leveraging techniques from explainable AI where insight is provided into a robot's decision-making to attempt to improve situational awareness. They introduced two complementary modalities of AR-based visual guidance for the human teammate, generated by a multi-agent search algorithm for uncertain environments: prescriptive guidance (the robot directly recommending actions), and descriptive guidance (the robot showcasing the state space information used in its decision-making). In a comparison of these modalities (as well as a combined interface), they found that the combination of prescriptive and descriptive guidance led to the highest perceived trust and interpretability, the highest task performance, and made human collaborators act more independently.

2.4.4 Medical

There are a number of applications of AR for improving human-robot collaboration in robot-assisted dental work as well as for robot-assisted surgery. [97] provide an extensive review of AR for robotic-assisted surgery, providing a comprehensive list of application paradigms: surgical guidance, interative surgery planning, port placement, advanced visualization of anatomy, supervised robot motion, sensory substitution, bedside assistance, and skill training. We will highlight some of the medical applications here to demonstrate a chronology, however for a full review of AR in robotic-assisted surgery, the reader should refer to Qian, Wu, DiMaio, **et al.** [97].

For performing dental work, Ito, Niwa, and Slocum [48] presents visual overlays in AR for a robot-assisted dental milling machine via teleoperation. Virtual objects are superimposed on physical objects, allowing the user to see the trajectory of the cutting tool path as well as a patient's internal bones.

For a situation requiring first aid, experts are often not at the site to provide treatment. It is specifically cases like these that Oyama, Watanabe, Mikado, **et al.** [98] attempts to address with a Remote Behavior Navigation System (RBNS). This system equips a person at the site of the emergency with a camera, microphone, and HMD, while a remote expert is able to view the camera feed and provide directions for care that are mimicked in the HMD virtually. The experiment challenges a participant to construct an arm sling using the RBNS, remotely guided by an expert.

The AR system presented in Filippeschi, Brizzi, Ruffaldi, **et al.** [99] is a complete system for remote palpation (examination by touch), in the case where a patient and a doctor are not collocated. Both visual and haptic feedback are provided to the doctor, and the patient is in view of an RGBD camera.

For assistance both before and during surgery, Adagolodjo, Trivisonne, Haouchine, **et al.** [100] develop an AR system for visualizing tumors and blood vessels around the surgery site. Approximate 3D pose information is obtained from 2D silhouettes, proving this method potentially useful for planning surgical operations. Similarly, in Zevallos, Rangaprasad, Salman, **et al.** [101], AR is used to show the shape and location of tumors by visually overlaying that information onto the actual organ, in an effort to assist surgeons. In this example the surgeons use the da Vinci Research Kit (dVRK), a robotic surgery assistant. A system is presented to autonomously locate the tumor, provide stiffness and related information about the tumor, and then overlay the information on a model of the affected organ for display to the user. Another application for surgery is from Qian, Deguet, Wang, **et al.** [34], where the First Assistant is provided with a HoloLens that is equipped to aid them with instrument insertion and tool manipulation while using the da Vinci robotic surgery assistant. Experimental results show potential improvement in efficiency, safety, and hand-eye coordination.

Elsdon and Demiris [30] use a HoloLens and a "spray robot" for dosed application of topical medication. Because sprayed dosage is difficult to visualize, the density is visualized virtually, and the Actuated Spray Robot is enabled with three different modes: manual (user must pull trigger and move sprayer), semi-automatic (trigger is actuated automatically but user must move the spray head), and autonomous (both the trigger and head articulation are automated). A more even density (greater accuracy) is achieved with both semi-automatic and automatic modes than with manual spraying, although manual was fastest. The experimenters speculate that because both of the automatic modes do not allow mistakes to be made, participants may tend towards perfection in those modes, increasing the time spent on the task. This technology could also be applicable in manufacturing, for paint and other coatings requiring a spray application.

2.4.5 Space

Space applications pose challenging problems, especially as the work sites reach farther and farther from earth. Any teleoperation must account for the time delays imposed by these long communication distances, a problem explored deeply by [102]. Xia, Léonard, Deguet, **et al.** [103] attempt to work within these constraints by using augmented reality to help simulate the time delay for a remote operator. Via AR, different virtual fixtures are tested to aid the operator, both with and without a time delay. Use of virtual line fixtures is the best option, with or without the delay, while using virtual planes decreases the task time to less than 1/3 of the unassisted task with a time delay. The design of this experiment, while in this case is applied to satellite repair, is derived from medical applications, and could have applications in this field as well, especially as it relates to medical care *during* space travel.

Somewhat surprisingly, literature on AR for HRC in space applications seems few and far between. Furthermore, most of the found literature is for remote teleoperation rather than collocation. We speculate that this could be due to a combination of factors. Most importantly, currently humans are only present in space in low Earth orbit, on the International Space Station or on brief launches in relatively small spacecraft. While some robots exist in these locations, the opportunities for incorporating AR into their use have been sparse. Furthermore, due to the time delay in communicating with remote robotic spacecraft and rovers, such as the Mars Exploration Rovers (Spirit and Opportunity) or the Mars Science Laboratory (Curiosity) prohibits convenient real-time HRC. Thus, more of the research related to these kinds of collaboration feature virtual reality or augmented virtuality instead. With upcoming missions due to land humans on the moon, and eventually on Mars, this is an area rich for future research.

Virtual, augmented, and mixed reality (VAMR) will be key components of HRI and autonomy as we head into the future of space exploration. While prior work applying VAMR to HRI is generally done in close proximity and without a significant time delay [20], [25], [29], [33], [62], with appropriate focus, VAMR can be adequately developed to assist both proximal and remote HRI. VAMR currently provides nearby operators with supplemental information and visual aids, increased situational awareness, and additional functionality and modes of communication [45], [62], [112]. These same tools can be applied for operational uses in both a collaborative Martian task or a more remote task on Europa. During a Mars EVA, a human-robot team might be exploring an area together. The human, aided by navigation cues and object recognition in their AR heads-up display, can guide the rover or quadcopter to conduct further investigation. In a cooperative construction task such as the one described above, information can be displayed in AR to provide insight into the robot's decisions to improve situational awareness, and visual aids might provide a shared mental model. Alternately, a remote operator planning a traverse on Europa could be experiencing the setting of the robot via VR, enabling them to program and execute commands using waypoints placed from a first person or overhead perspective. Planned paths can be previewed in VR with potential issues highlighted in the user's field of view. This experience also increases the operator's situational awareness of the robot's environment. Meanwhile, repeated use and training can allow the human to understand the autonomy decisions and capabilities from the visual perspective of the robot. In both situations, the use of VAMR provides additional safety for both the humans and robots.

2.4.6 Safety and Ownership of Space

The collaboration problem of indicating to humans whether a space is safe to traverse, whether space is "owned" by the robot, or whether it is otherwise occupied or available has been explored in a number of different studies. As mentioned above in Section 2.3.1, work in Bolano, Juelg, Roennau, **et al.** [47] displays to users the intended goal locations, paths, and swept volumes of the robot and its end effector. The technology in Sprute, Tönnies, and König [41] provides a human with the ability to restrict a robot's workspace by drawing on a tablet in AR. In Makris, Karagiannis, Koukas, **et al.** [73], shaded rectangular prisms in a human's AR HMD denote the "safety volume" in green and the "robot's working area" in red. Alternately, in Frank, Moorhead, and Kapila [40], red shaded areas of the working plane indicate prohibited regions for the robot, and green shaded areas indicate allowable regions that the robot can reach. Puljiz, Krebs, Bösing, **et al.** [86] also highlight the ability to denote safety zones using their HMD-based mapping and interaction methods in a robot work cell in a manufacturing environment. New work in spatial ownership during collocated activities also shows that AR-delivered visualizations alone are insufficient for achieving human compliance with robot instructions, even in a high risk environment when humans are in close proximity to potentially dangerous airborne robots [3].

Notably, the use of green and red seems mostly dependent on whether the human is teleoperating, programming, or otherwise controlling the robot (in which case green indicates areas they are allowed to move the robot into), or whether they are performing a task in parallel (in which case green indicates areas where they are safe from the robot).

2.4.7 Other Applications

While somewhat unconventional, the following applications provide unique and creative perspectives on the possibilities for implementing AR for HRC. These researchers are trying to push people's boundaries on what makes for a good AR/HRC combination. We included these unconventional perspectives with the intent to inspire future work envisioning such systems. These works ask questions like, "How can we make this something that might be useful every day?" and, "What do people think about incorporating robots and AR into their daily activities?"

In Ro, Byun, Kim, **et al.** [104], a robot is presented as a museum docent that uses projection-based AR to share information with human visitors. Applications for this technology might also expand past museums to malls and city streets, or even classrooms.

Mavridis and Hanson [105] designed the IbnSina (Avicenna) theatre installation to integrate humans and technology, and to provide a place for art, research, and education to come together. The stage is outfitted with sensors and is occupied by a humanoid robot along with humans. Though not yet fully implemented, the theater is intended to be interactive, and is to be equipped with a screen, lights, and audio and video systems, enabling holograms and interaction.

Anticipating future restaurant applications, Pereira, Carter, Leite, **et al.** [106] present a fast food robot waiter system in a wizard-of-oz study. Participants in a within-subjects study teleoperate the robot either solo or with a partner, using a headset and joysticks.

Omidshafiei, Agha-Mohammadi, Chen, **et al.** [107] outline the usefulness of AR when prototyping and testing algorithms. By combining physical *and* virtual robots in an augmented environment via the use of projection AR, motion capture, and cameras, different systems can be tested and evaluated in full view of the researchers, and without the risks involved in deploying them in the outside world.

Another nascent research area for AR-based HRC is Socially Assistive Robot tutoring, as in Mahajan, Groechel, Pakkar, **et al.** [108]. In this study, the researchers assess the use of common 2D usability metrics, such as *performance*, *manipulation time*, and *gaze*, and their correlation with usability scores from the System Usability Scale (SUS) survey. During an AR-assisted programming task, they find a positive correlation of usability with gaze, but not with manipulation time or performance.

2.5 Evaluation Strategies and Methods

In general, we are all working towards developing something "better." What we mean by "better," however, can have vastly different definitions based on the context and the intent. Better could be faster, more efficient, more directly, safer, with higher fluency, with greater situational awareness, or many other possibilities. In order to evaluate whether something is better, both objective and subjective measures can be made via multiple kinds of evaluations. These evaluations and measures are the subject of this section.

Because there are many aspects to evaluation, here we take a few different approaches. First, we highlight some instruments and questionnaires that have been used in evaluating AR for HRC. Then we discuss the choice to conduct extensive user studies, pilot testing, or only proof-of-concept testing, and the value of each of these options, as well as considerations for recruiting participants.

2.5.1 Instruments, Questionnaires, and Techniques

Summary of Instruments and Techniques for Evaluation

Instrument/Technique	Reference(s)
NASA Task Load Index (TLX)	[113]
Godspeed Questionnaire Series (GQS)	[114], [115]
User Experience Questionnaire (UEQ)	[116]
System Usability Scale (SUS)	[117]
Situational Awareness Evaluation	[118]
Task-Specific Evaluations	[119], [120]
Comprehensive Evaluation Designs	[121]–[123]

Table 2.2: This table summarizes the instruments, questionnaires, and techniques elaborated on in Section 2.5.1, along with the reference(s) applicable to each.

2.5.1.1 NASA Task Load Index (TLX)

Use of the NASA Task Load Index or NASA TLX instrument [113] is perhaps one of the most widespread in assessing AR for human-robot collaboration [30], [38], [40], [42], [46], [58], [74]. The NASA TLX assesses work load on six scales [113] and was originated by Hart and Staveland in 1988 [124]. The six scales are Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. The instrument is now available in both paper-and-pencil as well as mobile app format [113], making it very easy for the experimenter to deploy and for the subject to use.

2.5.1.2 Godspeed Questionnaire Series (GQS)

The Godspeed Questionnaire Series [114], [115] was developed by Bartneck et al. in 2009 as a way to measure "anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots". Each of these 5 areas contain 3-6 Likert-type scales on which to rate the robot. This questionnaire was used to measure "perception of an artificial embodied agent" in Rotsidis, Theodorou, Bryson, **et al.** [43], while in Williams, Bussing, Cabrol, **et al.** [63] only the *Likability* section was utilized.

2.5.1.3 User Experience Questionnaire (UEQ)

Both Bambušek, Materna, Kapinus, **et al.** [58] and Kapinus, Beran, Materna, **et al.** [75] utilized the User Experience Questionnaire [116], or UEQ, as part of the evaluation. The UEQ is a 26-item assessment; each item is ranked on a 7-point scale. The results provide a rating of the product being evaluated on 6 separate scales: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty.

2.5.1.4 System Usability Scale (SUS)

Measuring usability with the SUS is a method of quantifying a somewhat qualitative element of a design or technology. One measure of usability that a number of studies [46], [58], [78], [89], [108] utilize is the System Usability Scale or SUS [117]. The SUS consists of 10 statements that users can rank on a scale of 1 to 5, from *strongly disagree* to *strongly agree*). Example statements include "I think that I would like to use this system frequently" and "I found the system very cumbersome to use". To attain the total SUS score, for all odd numbered responses subtract 1, and for all even numbered responses subtract the response from 5. Add these scores together, then multiply the total by 2.5. This provides a score in the range of 0 to 100.

2.5.1.5 Situational Awareness Evaluation

A common claim is that AR lends itself to increasing the user's situational awareness, or SA. Many papers in this survey claimed to evaluate situational awareness [25], [40], [88], [97], [125]–[127], but few actually had a way to evaluate this [31], [33], [62], [87]. Endsley [118] defines situation awareness as "the pilot's internal model of the world around him [sic] at any point in time," what roboticists might refer to as a *mental model*. Specifically, a version of the Situational Awareness Global Assessment Technique (SAGAT) developed by Endsley [118] is used in Srinivasan and Schilling [125]. The SAGAT was developed in 1988 (interestingly, this also coincides with the original publication of the NASA TLX) to assess aircraft designs for pilots' situational awareness. Scholtz et al. adapted the SAGAT in 2004 for (semi-)autonomous vehicles ("robotic vehicles") and human-robot interaction, specifically the "supervisory role" that humans play in this situation [128], [129]. In the original SAGAT, the experiment is paused at various points throughout the study, and during these pauses the pilot/subject is asked a series of questions that are intended to assess their awareness of aspects of the current situation. The evaluation is given via computer to allow for randomized questions as well as rapid response inputs. A composite score is acquired based on the total response results. It is important to note that SAGAT is a *technique* and not a specific instrument or questionnaire. The particular questions asked during each pause or interruption are entirely dependent on the environment in which SA is being evaluated.

2.5.1.6 Task-Specific Evaluations

When conducting a user study, the researchers should conduct a thorough search to discover existing instruments for their technology's particular use case.

For example, in testing the functionality of an AR design to be used by robotic wheelchair operators, Zolotas, Elsdon, and Demiris [26] choose skills from the Wheelchair Skills Test, version 4.2 [119], [120]. The most current version of this manual is now version 5.1 [130], and it contains the specifics of the Wheelchair Skills Test, or WST, with individual skills, a questionnaire (WST-Q), and training. Examples of the skills assessed include *turn while moving forwards (90°), turn while moving backwards (90°)*, and *gets over threshold (2cm)*. Because there is an established test and instrument for these kinds of skills, it follows that the WST and WST-Q would be used to evaluate an AR system intended to assist robotic wheelchair users.

2.5.1.7 Comprehensive Evaluation Designs

Experiments in Kalpagam Ganesan, Rathore, Ross, **et al.** [45] utilize "questionnaire items...inspired and adopted from Hoffman [121] [since updated in Hoffman [2]], Gombolay, Gutierrez, Clarke, **et al.** [122], and Dragan, Bauman, Forlizzi, **et al.** [123]." Here we discuss why these three works present ideal fodder for comprehensive questionnaires.

In Hoffman [2], Hoffman defines fluency in HRI and then presents metrics for measuring fluency. In defining *fluency*, he states that, "when humans collaborate on a shared activity, and especially when they are accustomed to the task and to each other, they can reach a high level of coordination, resulting in a well-synchronized meshing of their actions. Their timing is precise and efficient, they alter their plans and actions appropriately and dynamically, and this behavior emerges often without exchanging much verbal information. We denote this quality of interaction the fluency of the shared activity." Hoffman also clarifies that fluency is distinct from efficiency, and that *people can perceive increased fluency even without improvement in efficiency.* These fluency measures include both objective (for example, percentage of total time that both human and robot act concurrently) and subjective metrics (for example, scale ratings of trust and improvement).

Both Gombolay, Gutierrez, Clarke, **et al.** [122] and Dragan, Bauman, Forlizzi, **et al.** [123] actually draw substantially from the measures presented in Hoffman [2]. [122] choose to use 13 questionnaire items from the subjective metrics in Hoffman [121] and augment this list with 8 of their own "Additional Measures of Team Fluency," focused on the human's satisfaction with the teamwork. [123] use both objective and subjective measures from Hoffman [2], and add items related to closeness, predictability, and legibility.

We recognize that none of the studies that Kalpagam Ganesan, Rathore, Ross, **et al.** [45]

draws from are necessarily related to the use of *augmented reality* for human-robot collaboration. However, the relevance and appropriateness is apparent, and can easily be used in combination with other metrics specific to AR.

2.5.2 The Choice to Conduct User/Usability Testing

Three main themes in testing and evaluation emerge from the papers reviewed. (1) **Pilot testing** provides a way to verify that research, technology, or evaluation is headed in the right direction, or to determine certain specifics about a subsequent evaluation. (2) **Proof of concept experiments** or prototypes can demonstrate that a particular technology can in fact be implemented, and might also highlight additional directions to take the research. (3) **User or usability testing** provides the researchers with feedback and data on their current designs; the better the participant pool (again, note that "better" is a loaded word here), the more trust they can typically have in their results. We look more deeply at each of these three themes in this section.

2.5.2.1 Pilot Testing as Verification

Some studies use a pilot test to then inform a larger scale test that is also described in the same paper. In Qian, Deguet, Wang, **et al.** [34], where the authors present a form of AR to assist a surgeon's First Assistant with the da Vinci robotic manipulator, they first perform a pilot test with 3 surgeons. After this initial evaluation, and using feedback from the pilot subjects, they then conduct an n=20 user study. [74] briefly mention an initial pilot study to evaluate whether pointing and head gaze were natural modes of selection for a user, before explaining their more thorough n=16 user study. In Sibirtseva, Kontogiorgos, Nykvist, **et al.** [57], a human-human pilot study is conducted (n=10), where data is collected on the vocabulary used to describe Lego objects between human partners. Informed by this pilot, the authors decide to resort to a wizarded system for the natural language processing portion of their experimental setup.

Alternately, other studies *only* present on a pilot test, then address how this test might inform future, larger scale testing. [131] report on their pilot study (n=10) that requires users to complete

2 tasks in 2 different conditions: the experimental condition of a "proposed AR-robotic interface" and a gamepad. These authors then proceed to discuss a case study, where the technology is applied to the process of carbon-fiber-reinforced-polymer production, and then pilot tested on 1 user. To evaluate the design of an AR HMD for wheelchair users, [26] run a between-subjects pilot test on 16 participants who must navigate a route 4 separate times, either with or without the AR visual assistance. All of the results can inform future iterations of the design. In Yew, Ong, and Nee [25], a pilot test is presented using their prototype, to show that combining virtual objects with in situ spaces can function for teleoperation of robots. Tasks are completed by the novice users (n=5) in a short amount of time, setting the stage for future evaluations and also revealing areas for improvement of the design (tracking sensors and algorithms, depth sensors for unforeseen hazards).

2.5.2.2 Usability Testing

Throughout this paper, there have been examples of numerous studies that conduct full usability or user testing. Some highly cited examples include Walker, Hedayati, Lee, **et al.** [18], Hedayati, Walker, and Szafir [33], and Chakraborti, Sreedharan, Kulkarni, **et al.** [60]. Commonalities among these experiments include a relatively high number of participants and a thoroughly and intentionally designed study. In all of these examples, participants take part in the study in person. Another option is to perform testing using Amazon Mechanical Turk (MTurk) users who view videos or simulations of the system. By using MTurk, the number of subjects can often be expanded, however limitations include the mode of interaction and the kinds of participants.

2.5.2.3 Proof of Concept Experiments

The two kinds of evaluation presented in Sections 2.5.2.1 and 2.5.2.2 are both intended to gather objective data (for example, how long a task takes to complete or where there is overlap in the duties of the human and the robot) as well as subjective data (for example, whether the human user understood a command or preferred a certain type of interface). Meanwhile, other experiments published show that a technology can indeed be implemented in a certain way, with the intent to solve a particular problem. One example of this kind of experiment is in Reardon, Lee, Rogers, **et al.** [62]. In this work, the authors thoroughly document how they successfully implemented an AR display for use in assisting a human user while they collaboratively explored a potentially dangerous space with a ground-based robot. They combine an understanding of cooperative exploration with complete integration of the robot's and human's points of view, and augment this with additional data provided to the human by the robot. In the experiments described, the system successfully performs all necessary tasks.

Other examples of a proof of concept study include a generalized AR system that is laid out for human operators working with assembly line robots in automotive manufacturing [73], an AR/VR system in collaboration with a ROV designed to enable virtual SCUBA diving [50], virtual drag-and-drop programming of a robot arm for a pick-and-place task [28], robotic-assisted masking of areas for mechanical repairs [132], a system for AR-enabled online programming of industrial robots including motion and hand gesture tracking [36], an architecture for implementing AR for programming robots using multiple modalities in industrial settings [24], and the use of built-in mapping functionality in a HoloLens to establish the working environment for a robot arm in a work cell [86].

2.5.2.4 Choosing the Type of Evaluation to Conduct

How does one choose the right kind of evaluation for a particular technology or study? Elements to consider include: (a) how far along the technology is in its development, (b) how many test subjects it would take to validate or evaluate the design, (c) whether the technology is safe for human subjects, (d) what research questions are being asked. Sometimes a pilot study may be warranted to obtain additional details before proceeding. In other cases it is only the technology that needs to be showcased, and extensive user testing is not necessary. If the researchers are attempting to show increased usability, safety, or fluency, a full scale human subjects experiment will be necessary. We recommend starting by examining the goals of the evaluation, for example framing it in terms of one of the previous three sections (pilot testing, usability testing, or proof of concept). From there, similar studies can be referenced that have comparable intents. Informed by this survey and prior work, the researcher can choose appropriate instruments or evaluation techniques for their own purposes.

2.5.2.5 Recruiting Participants for Human Subjects Studies

We would also like to address the issue of recruiting participants for user studies. There are multiple factors to consider, all related to diversity in the participant pool, which we enumerate here.

- **Diversity in experience.** Novice participants are often recruited local university student population out of convenience. Researchers should consider whether recruiting experienced or trained participants (who might be experts or professionals in the tasks being performed) might benefit their study.

- **Diversity in age.** Again, if the participants are mostly recruited from one age group, such as university undergraduates or employees of one group at a company, their prior experiences may prove to be somewhat uniform. As technology continues to advance rapidly, participants of different ages will inevitably have varied technological literacy. Researchers should consider the impact this might have on their results and what they are seeking to learn from the study.

- **Diversity in gender, race, and ethnicity.** User study participants should be recruited to reflect the population as a whole (see Palmer and Burchard [133]). As with the prior items in this list, participant populations that are not representative can affect the usefulness of the results.

Most importantly, researchers must recognize in any publications the shortcomings of a participant population. Demographic and other relevant information about participants can help clarify what these gaps might be and allow for critical reflection on whether this could have affected any results.

2.6 Classifying XR for HRI

To help advance characterizing VAM-HRI systems, we introduce a **Tool for Organizing Key Characteristics of VAM-HRI Systems (TOKCS)** [15]. TOKCS builds off work from the Interaction Cube [134], discretizing its continuous scales and adding new key characteristics for classification. The tool is applied to the 10 workshop papers from the 4^{th} International Workshop on VAM-HRI [12] to validate its usefulness within the growing subfield. These classifications help inform current and future trends found within the workshop.

The Interaction Cube uses three dimensions to characterize VAM-HRI work: the 2D Plane of Interaction to represent interactive design elements and the 1D Reality-Virtuality Continuum from Milgram to characterize the environment. The first two dimensions of the Interaction Cube are defined by the *Plane of Interaction*, which captures both (1) the opportunities to view into the robot's internal model, and (2) the degree of control the human has over the internal model. The third axis of the Reality-Virtuality Interaction Cube illustrates where an MRIDE falls on the Reality-Virtuality Continuum [135].

TOKCS consists of characterizing VAM-HRI systems with: Anchor Location {User, Env, Robot}, Perceived Manipulability {User, Robot, None}, Increases Expressivity of View (EV) {0,1}, Increases Flexibility of Controller (FC) {0,1}, Increases Complexity of Model (CM) {0,1}, Milgram Continuum {AR, AV, VR}, Software Description, and Hardware Description.

2.7 Future Work

The field of augmented reality for human-robot collaboration is vast. One can examine the suitability of various AR technologies for an HRC task, the design of the AR interfaces, the user experience, the comfort, and the safety. We can ask questions about what humans are capable of, how the human and the robot can work together or separately, how much the human should be asked to do, or how they should be asked to do it. Alternately, we can ask questions about what the robot can do, how the robot should be instructed or programmed, and what levels of tasks it can perform. At a system level we can design systems that seamlessly integrate a human, robot, and AR device; we can examine behaviors of systems in all kinds of environments, indoors and outdoors; we can evaluate how well the systems function either remotely or in situ. The 2020 Robotics Roadmap [136] assembled by a consortium of universities in the US lays out some specific current challenges for human-robot interaction, including accessible platforms, datasets, and evaluation. All of the works presented here take various perspectives on these questions and more. However, as with all research areas there is still much to explore. Here we will touch upon a few key areas that are calling for innovation and improvement.

In many ways, the field will continue to evolve with the maturation of augmented reality technology, including next generations of head-mounted displays, improved handheld AR, and possibly even innovations to projection-based AR. As recounted in Puljiz, Stöhr, Riesterer, **et al.** [29], issues with segmentation demonstrate the need for improvement in AR capabilities with regard to skin color, limb, and gesture recognition. AR must be able to work in all kinds of environments regardless of lighting, background, or the user's skin color in order to be effective. Furthermore, in Kästner and Lambrecht [32] the main limitations are from constant visualization of real-time data, especially the laser scan data for position and obstacle tracking. These difficulties demonstrate the current processor and visualization limitations in AR technology.

AR technology has also been described as bulky [45], cumbersome [137], and having a limited field of view [26], [34], [57], [138], [139]. All of these issues present opportunities for improvement of the AR technology itself.

Collaboration of HRI researchers with those developing cutting edge user interfaces should also be emphasized. In order to obtain accurate and meaningful results from user studies, AR interfaces must utilize established principles of design for accessibility and functionality. In Stadler, Kain, Giuliani, **et al.** [38], the authors suspected that because of an excess of detailed information provided through AR, users actually took more time to complete a task that should have been faster with the help of the AR display. Questions such as *What is the appropriate level of information to provide to someone performing an AR-assisted task?* could be asked of a UI designer and incorporated into future work.

2.7.1 Robots and Systems Designed to Be Collaborative

The works included in this review typically utilize one robot (ground-based, robotic arm, aerial, underwater, or humanoid) in collaboration with one human. The robots are designed for a variety of purposes - to be universal manipulators, drive over smooth or rough terrain, or easily navigate in a three-dimensional space. But not all of these robots are designed expressly for the purpose of working in close collaboration with humans. Some were chosen based on their ease of manipulation in a programming-by-demonstration task or their safety features. However, what happens when we *first* take into account the possibility that a human might be working in close proximity? What kinds of features can we innovate to ensure the person's safety as well as ensure that the robot completes its task? How might this robot behave? And what might this collaborative environment look like in different environments?

2.7.2 Humans as Compliant Teammates

Much work exists that explores the role of the human as the director, manager, or overall controller. But what if we turned this idea on its head and made the human a vital component on a robot-driven team? What if AR was utilized to direct one or more humans in a collaborative task with one or more robots? What if we were able to easily expand past the currently typical robot-human dyad, which the vast majority of the works surveyed here involved?

Furthermore, we are continuing to think of these as human-robot *teams*. The goal is not to replace human workers altogether, but to utilize the strengths and intelligences of both humans and robots to increase productivity and efficiency. How can we make both humans and robots more productive by teaming them together? As Reardon, Lee, Rogers, **et al.** [62] point out, we want to "influence the human's model of the robot's knowledge and behavior, and shape the human's performance. In this way, we treat the human and robot teammates as peer members of the cooperative team, and seek to influence each through information communication."

2.7.3 Evaluation

In Section 2.5 we summarize different methods of evaluating a technology and measuring improvements. However, it is also obvious how much room for innovation there is in this particular area. There are very few standardized, validated, and widely used instruments. Pick-and-place and other manufacturing-related tasks are also prevalent in the literature, yet few evaluation methods are alike, making it difficult to compare across different studies. Greater collaboration among researchers could yield some semi-universally accepted evaluations for typical AR for HRC tasks, such as teleoperation (both remote and in situ), aerial robot piloting and communication, or pick-and-place tasks.

2.8 Conclusion

We are thinking ahead to a future when robots will be able to plan and execute even more efficiently and when augmented reality is an unobtrusive and fluid method of communication. Augmented reality will only continue to mature into a more accessible technology, and its role in human-robot collaboration can become much more impactful and relevant to HRI. Specifically, with visual and graphical communication delivered via AR, we can provide a human teammate with information about spatial ownership. These developments lead us to ask whether AR can be used to truly help keep humans safe when working around autonomous robots. We explore this deeply in Chapter 3.

Chapter 3

Human Non-Compliance with Robot Spatial Ownership Communicated via Augmented Reality

3.1 Introduction

Due to a confluence of technological availability and utility, humans and robots are increasingly operating in close proximity to each other. The current state of safety in human-robot collaborative and cooperative collocated tasks generally revolves around protecting the human from any contact with the robot, using physical barriers and sensors to pause robot operation in the vicinity of humans. This is oft realized as robots installed within physical cages or within fences in a manufacturing environment, or as ground robots in a well-structured warehouse environment that stop when a human approaches wearing specially instrumented clothing. While effective at preventing negative interactions, these approaches tend to be inefficient and cause frustration.

Research on increasing predictability and interpretability of quadcopter robots by collocated humans [18], [41], [140], [141], in addition to work that predicts human movement [142], tends to assume that a robot should always defer or conform to human preferences independent of the rationale behind them. However, the practical alternative of expecting the human to conform to the robot's movements or demands is less explored. With increased deployment of robots in established processes within warehouse, manufacturing, and even space environments, we must find safe, efficient, and robust ways of collaborating with them.

This work surfaces insights about human compliance and non-compliance with robot instructions for spatial ownership as delivered via augmented reality in a collocated environment with

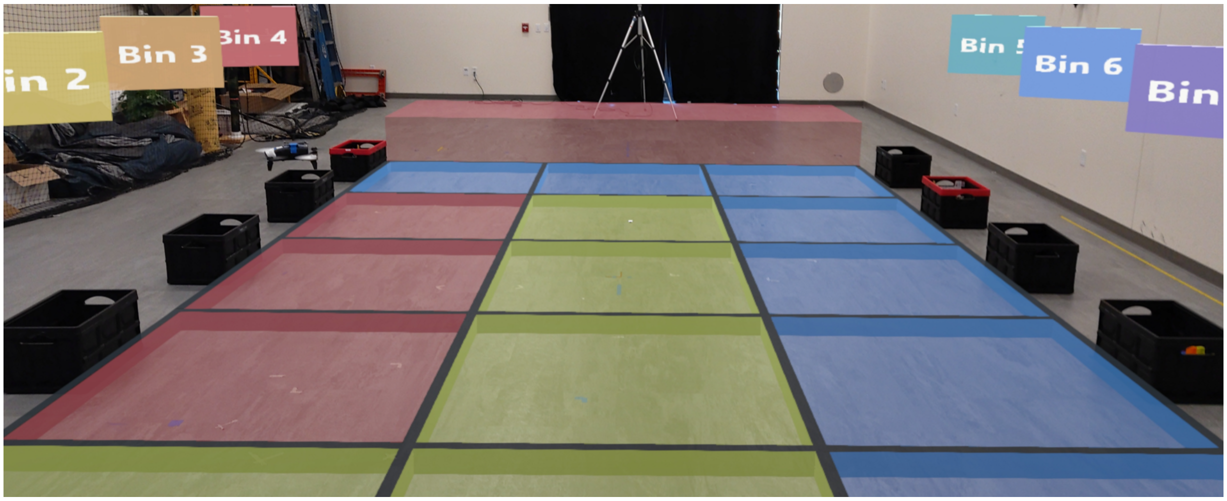


Figure 3.1: View through the HoloLens from Home. Eight bins contain task components, a 3x5 grid indicates spatial ownership, and bin labels are selected to request access. Shown is a path to Bin 4.

Figure 3.2: Looking towards Home in the ARHMD during the trap scenario (no return path) in the Shared Space condition.

important safety implications. These insights are gathered from an experiment where human and robotic agents held ownership over different areas of a warehouse floor. We designed and implemented the FENCES (Facilitation of Efficient Nonverbal Collocated Environment Safety) System to enable this interaction. FENCES enables a user to request permission from an autonomous robot to traverse the work floor to reach bins containing parts needed for an assembly task. The robot, an autonomous free-flying quadcopter that is conducting an inventory task, gives permission by giving the human temporary ownership of parts of the floor indicated by hologram coloration (see Figure 3.1).

We investigated user behavior and compliance with respect to the FENCES system through an Institutional Review Board-approved, between-subjects study with two conditions: (1) a shared-space condition where the human and robot occupied the floor concurrently, and (2) a turn-taking condition where the human and robot performed their tasks sequentially, with only one of them allowed on the work floor at a time. The main contributions of this work are our findings surrounding human compliance and the justifications they provide for non-compliance and the subsequent identification of critical design considerations for future AR-based safety systems to incorporate, with implications for safety, trust, and cognitive load.

3.2 Related Work

The FENCES system and the experimental design in this work are based on insights synthesized from collections of research within the multiple interconnected themes of communication, safety, augmented reality, and human-robot interaction, expanded upon in the subsections that follow.

Communication of Information in AR. McIntire et al. [143] find that stereoscopic 3D displays have equal or superior information communication performance as compared to non-stereo (2D) displays the majority of the time. Augmented reality (one form of stereoscopic 3D display) is a preferable option due to its dynamic visualization capabilities, non-obstruction of the visual field, and relative ease of use. Szafir and Szafir [144] indicate that most past research on human-robot interface design has centered around situational awareness and user control. While our system provides situational awareness in terms of spatial ownership, we look beyond control and towards back-and-forth communication between the human and the robot.

AR for Human-Robot Communication. A rich corpus of work on use cases and experiments exists regarding using AR for human-robot communication [18], [33], [42], [61], [63], [71], [74], [80], [89], [131], [140], [141], [145]. Many systems are designed to improve communication **from the robot to the human**, such as providing insight into motion intent [18], [71], [140], assistive control predictability and legibility [89], aiding teleoperation [33], improving control handovers for autonomous vehicles [145], and using AR-assisted robot gestures [63]. Other systems exist that facilitate communication **from the human to the robot**, including programming or otherwise adjusting the system [42], [74], [80], [131], teleoperation [33], providing boundaries to the robot [41], [61], or functioning as a team [146], [147]. While our work builds on this growing body of research, we specifically address human ***compliance*** with a communicative system as it relates to safety.

AR and Safety. AR is increasingly used to improve worker safety in a variety of environments [148], [149]. Tatić and Tešić [150] presented a case study using AR to improve safety in an industrial environment by providing virtual safety instructions and other information. AR-equipped hard hats are also increasing in prevalence, indicating there is growing acceptance of using AR in high-risk environments [151], [152]. Our work leverages these findings and techniques in ***spaces containing humans and robots***.

AR for Human Safety in Shared Spaces with Robots. A system from Choi et al. [153] provided safety signals in the form of a green, yellow, or red dot for low, medium, and high risk of danger in the corner of the user's field of view. Makris et al. [73] also shaded regions of the workspace in red to denote the robot's space or green to indicate the operator's safe working area. In practice, for our system we found that users had difficulty distinguishing between yellow and green holograms, leading to our use of blue instead of green, but maintaining the overall principle of using color to denote ownership or imply safety.

Some primary applications for our findings include manufacturing and fulfillment centers. There are indications that humans working in close proximity to robots at Amazon Fulfillment

Centers might alter their workflow to accommodate or support the work of their robot teammates [154], prompting the authors to ask how AR can further facilitate these human-robot teams. Amazon has already initiated work on this front, as evidenced by the existence of a patent on an AR display for fulfillment center workers [155], [156].

In this work, we utilize augmented reality to provide both a communication modality and spatial ownership information for a person working collocated with an aerial robot and draw conclusions related to human compliance and safety.

3.3 The FENCES System

The FENCES system includes a Microsoft HoloLens 2 augmented reality head-mounted display (ARHMD), a Parrot Bebop 2 quadcopter robot, a Vicon Tracker motion capture camera system for tracking the robot and the user, and a computer performing sensor fusion, state management, and robot control. In the component descriptions below, the term “user” refers to the human participant.

FENCES was designed as a test bed for analyzing human behavior while interacting with AR and a collocated robot. Within the system, a user can request permission to traverse a controlled space in order to reach a specific goal location. Through the ARHMD, the user can see who has ownership of the spaces on the floor: the robot, themselves, or no one.

3.3.1 Mobile Robot: Parrot Bebop 2

The Parrot Bebop 2 quadcopter robot is an agile aerial vehicle approximately 330mm wide and 90mm tall. In our experimental setup, it flies 1 meter over the ground and emits a fairly loud noise ($\sim$70dBA at 1m) when airborne. Its blades are unprotected by guards, increasing its imposition on participant safety.

3.3.2 Microsoft HoloLens 2 ARHMD and User Interface

The Microsoft HoloLens 2 is capable of projecting images and text in the wearer's field of view. The user interface was designed in Unity [157] and consists of the following features, some of which can also be seen in Figures 3.1 and 3.2: (1) A large 3-by-5 grid on the floor, with the 8 bins and table serving as boundaries. (2) The 1.5 x 1.5 meter grid squares are colored red, yellow, or blue, depending on whether they are "owned" by the robot, no one, or the human, respectively. (3) Billboards above each bin are labeled with a corresponding number and always face the user. They can be selected using a HoloLens "air tap" to indicate a user request. Audio feedback is provided when a bin/billboard is selected ("Bin [number] selected."). (4) A "Home" billboard hovers above and behind the home base table. (5) Text confirming completion appears when the experiment has ended. The ARHMD is the sole mode of communication between the user and the system. The user initiates a request to approach a bin by selecting its billboard, and the system may give permission to traverse the floor, indicated by shading the grid squares in blue that the user is permitted to enter.

3.3.3 Experiment Manager and Experimenter Interface

All of the robot goal locations, floor color configurations (and thus user access routes), and anticipated bin selections are *predetermined* by the experimenters and implemented as sequentially reachable states in the system. The states have transition criteria based on specific conditions being met: user location, robot location, and bin request.

3.4 Experiment Design

We designed this IRB-approved experiment (n=*20*) as a between-subjects study with two conditions. Participants were assigned pairwise randomly to conditions: odd numbered participants were randomly assigned a condition and the following even numbered participant received the opposite condition. Pairwise randomization is an unbiased assignment mechanism to ensure balanced cases when there is a guaranteed pair [158]. We recruited 22 participants, but two trials were discarded due to issues with the motion capture system. The participant population drew from students at our university and was 25% female, 5% nonbinary, and 70% male. On a scale from 1 ("Never interacted with") to 5 ("Extensive experience with"), average experience across participants was 3.1 for robots and 2.4 for AR.

We deployed FENCES in an experimental flight space lab arranged to replicate an assembly environment with eight distributed parts bins along east and west sides (Figure 3.1). The task space was approximately *8m x 12m*. A table for the user's workspace was at the south end, deemed "Home Base" for the human. Participants received an orientation at this table, which also contained the assembly workspace and instruction booklet. The experimenter and control equipment were behind protective netting to the west of the table.

After signing the consent form, participants read one page of instructions describing the experimental task. The activity involved constructing a small assembly with Mega Bloks according to a printed booklet of step-by-step instructions with words and photos (see Figure 3.3). They were instructed to collect the blocks from the bins in a strict order from the bins and told that they should only walk on the blue areas in the grid. While the yellow and red areas of the grid were functionally similar for the human (areas not to walk across), the red areas were owned by the robot, while the yellow areas were not assigned ownership. Participants wore the Microsoft HoloLens 2 ARHMD described in Section 3.3.2, which provided the interface for users to request permission from the robot to traverse the space and obtain access to particular parts bins (see Figures 3.1 and 3.2). Simultaneously, the quadcopter robot flew about the room, stopping at bins to simulate inventory checks.

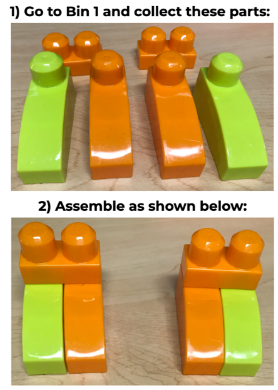
The two conditions were designated "Shared Space" (SS) and "Turn-Taking" (TT). In the **SS condition**, the participant and the robot were permitted to work in the grid area simultaneously, in non-overlapping regions of the space. The robot never returned to Robot Home, a red, robot-only location at the north end of the grid analogous to the human's "home base". The entire task took approximately 15 minutes to complete in the SS condition. In the **TT condition**, the participant alternated with the robot occupying the floor space; while the robot conducted its inventory route, the participant was required to stay in their respective home base, and while the participant was collecting items from a bin and traversing the grid, the robot hovered at Robot Home. After each inventory excursion, the quadcopter returned to Robot Home via the same general path by which it had departed. Since the robot and the participant were never on the grid at the same time, the duration for the entire task increased to approximately 30 minutes. In both scenarios, the "ownership" of the grid squares (robot, human, or neutral/unowned) was communicated to the participant using the virtual grid described in Section 3.3.2 and pictured in Figures 3.1 and 3.2.

These conditions were chosen to investigate behavior in two different yet equally relevant situations: one where the spatial ownership rationale was more recognizable (Shared Space) and one where the spatial ownership rationale and associated safety concerns were less obvious (Turn-Taking). Participants were not provided explicit explanations in either condition about why certain regions were permitted or prohibited, only what the colors denoted. Because we were investigating behavior with respect to the floor ownership as designated in AR, we do not compare their behavior to an AR-free condition. Further, without any indicator of spatial ownership or a significant deviation of the quadcopter's behavior, travel through the space would have been prohibitively unsafe for participants.

Immediately after the task ended, participants answered verbal questions about their experience in an interview with an experimenter. They were asked about their thoughts and behavior during the experiment, as well as whether they perceived any inefficiencies and whether they felt unsafe. Finally, they responded to a survey consisting of Likert (5-point scale) and free response questions.

3.4.1 Land Scenario

In both conditions, the robot landed on the workspace floor approximately 60% of the way through the experiment. This scenario was designed to reduce the perceived risk involved in the shared space condition, since the robot was not currently flying, potentially tempting the participant to disregard the floor ownership indicators and to return home via a more direct route. In the TT



(a) An example of the instructions provided. Each appeared on separate pages for clarity.

(b) The completed assembly of multicolored MegaBloks.

Figure 3.3: The task (a) instructions and (b) final assembly.

condition, this also served to allow the experimenter to quickly replace the robot's battery with minimal disruption to the experimental timing of quadcopter behaviors between conditions.

3.4.2 Trap Scenario

Partway through the experiment, the participant requests access to Bin 4 and access is granted (Figure 3.1). Once the participant arrives at Bin 4, only the grid squares along the northern edge remain blue while the rest of the workspace floor turns red, effectively eliminating their route back to Home Base (Figure 3.2). The system then begins a 60-second timer, after which the path back to Home Base will reappear. The quadcopter hovers adjacent to Bin 1 in the SS condition and hovers at the Robot Home in the TT condition.

3.4.3 Hypotheses

Through the system and experiment described above, we test the following hypotheses: **H1:** Participants will feel safer in TT than in SS due to the reduced proximity to the quadcopter. This will lead to increased deviations in TT, as participants will rely on potentially faulty reasoning (i.e., based only on priors and directly observable features) when determining whether to follow the system guidance. They will also spend more time on the grid in SS due to increased caution near the robot. **H2:** Longer or less direct routes will invoke more deviations from the blue path than shorter or more direct routes. Thus, the land scenario and trap scenario will also invoke deviations that participants will self-justify.

3.5 Results

3.5.1 Mixed Methods in HRI

For a model of mixed methods analysis, we consulted Veling and McGinn's [159] recent survey of 73 qualitative research papers in human-robot interaction, specifically the categories of insights-driven, design, and hypothesis-driven studies. There is a substantial history of prior work in HRI that use qualitative and mixed methods [160]–[162]. Using widely accepted qualitative methods we gathered data in semi-structured interviews as well as textual analyses [159], and coded the responses for repeated key words and themes.

3.5.2 Trap Scenario

A striking 25% of participants chose to walk through the red and yellow regions to return to Home, disregarding the instructions they had received at the start to only walk through blue regions. Three were in the TT condition while two were in the SS condition, showing similar rates of non-compliance regardless of robot proximity.

- "I knew I was faster than it, so [wherever] it was gonna go I was gonna get out of dodge before it could get there." (TT)

In fact, in one case it seemed that *because* a participant had high trust in the robot's consistency, they disobeyed the floor colors to return to Home Base.

- "I can see that it's safe, so [walked through the red]." (TT)

Eleven of the 20 participants became impatient or assumed a malfunction when the trap scenario began and selected the "Home" button as a solution; 7 participants considered requesting another bin to generate a path, such as one close to Home, or Bin 4 again (the bin where they were trapped); 2 participants admitted that they considered going around the experiment area, outside the grid entirely.

- "I did come close to wondering whether [to walk] around the outside because...nothing will be there..." (TT)

When asked why their path back to Home disappeared, participants generally thought that there was a software issue (n=7) or that the robot was claiming the area (n=9).

- "It seemed like there was a glitch so I broke the rule [and] went straight through." (SS)

However, there was no significant correlation between participants' reasoning about why the path disappeared and their decisions about what to do, suggesting that all of the reasons provided warrant consideration. Furthermore, we can see that when an autonomous system lags, users will not wait patiently; instead they desire ways to work around the lag.

3.5.3 Safety, Efficiency, and Trust

One of the most remarkable results from this study was that **all participants felt safe during the experiment, with the exact same distribution across both conditions**. Given the statement, "I felt safe throughout the exercise," all responses were either 4 or 5, with an average of 4.7 (see Table 3.1). Furthermore, 7 participants mentioned the word "safe" in the interview *before* they were asked whether anything felt unsafe about the experience. Two participants used the word "safe" in their response to the question, "Did you find anything inefficient about this process?" One participant (SS condition) believed the system was too safe:

- "It's overly safe...there's not enough risk involved." (SS)

When asked if they thought anything was inefficient about the system, of the 20 participants, 17 identified inefficiencies, while 3 did not. One TT participant described a SS environment that would be more efficient, but SS participants had suggestions as well:

Table 3.1: Summary responses to select survey items. Parentheses indicate SS responses. 1 = Strongly disagree, 5 = Strongly agree.

Statement	**1**	**2**	**3**	**4**	**5**
I felt safe throughout the exercise.	0	0	0	6(3)	14(7)
I deviated from the given path during the exercise.	15(8)	0	0	1(1)	4(1)
I felt informed throughout the exercise.	1	1(1)	4(1)	7(3)	7(5)

- “There were...times where there was a yellow part that didn’t belong to anyone, and it still made me go around.” (SS)

Participants also volunteered their thoughts about trust, sometimes combined with issues of safety and efficiency:

- “...I trusted the robot to stay in its red areas.” (TT)

- “I trusted it. I think it was very safe at the cost of efficiency, I’d be comfortable with less safety if possible.” (SS)

As presented in Table 3.2, participants in the SS condition felt that the robot was more fair than those in the TT condition, suggesting a willingness to sacrifice safety for a perception of fairness. As expected and required by the experiment design, participants were closer on average to the robot in the SS condition (3.7 m) than in the TT condition (4.4 m), with $p<0.0001$ (Figure 3.4). However, the self-reported feelings of safety showed identical data for the two conditions (Table 3.1), **suggesting that participants felt as safe nearly 3 meters from the robot as they did when it was waiting predictably in Robot Home**.

Participants in the SS condition responded statistically significantly more positively to the statement, “I thought the robot was fair,” (see Table 3.2), suggesting that the longer wait time in TT implied a level of unfairness.

- ‘The robot thought its priorities were more important.” (SS)

Table 3.2: Mean responses, by condition, to select survey items. *$p<0.05$

Statement	Shared Space	Turn-Taking
I thought the robot was fair.*	4.0	3.1
I liked the way I interacted with the AR device.*	4.7	3.8
I thought the robot was very responsive to my requests.	3.7	3.0
I thought the robot was intelligent.	3.4	2.7

Participants further personified the robot and the system in some of their interview responses:

- "Sometimes you...had to...wait a little bit for it [the robot] to realize, 'Wait, I don't need that square, I can give it up."' (SS)
- "It knew when I was on the field and when I wasn't." (TT)

We also noted how many times each participant checked the robot's position by looking at it while they were on the grid. Data shown in Figure 3.5 indicate with statistical significance that the higher they perceived its intelligence, the fewer location checks a participant made. Repeated checks for the robot suggest that the human is engaged in tracking the robot. As multitasking increases cognitive load [163], **this suggests that increasing the perception of intelligence can be a powerful way to reduce cognitive load.**

During the pre-experiment briefing, the experimenter interacted with participants on the south side of the table located at Home Base, facing the grid. However, 4 participants chose to work from the north side of the table with their backs to the robot and grid as they were constructing the assembly. One participant chose to work from the west end of the table. This behavior (working without view of the robot) is possibly another indicator of participant trust in the system.

A number of other interesting behaviors were observed. Despite being told to conduct the tasks in the order provided in the instruction booklet, one participant in the TT condition attempted to increase efficiency by gathering blocks from more than one bin per excursion, for example if the

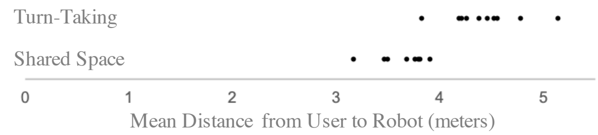


Figure 3.4: Mean distance from the robot sorted by condition. Across all participants, mean distance in SS = 3.7 meters, while mean distance for TT = 4.4 meters, $p<0.0001$.

following bin was also in the given path, as well as by trying to select future bins while he was on the grid. Other participants tried to anticipate what side of the table the path would start from, waiting on their predicted side, though frequently the path that appeared started from the opposite side as predicted.

3.5.4 AR Interface

Despite having the same user interface across both conditions, participants in the SS condition responded statistically significantly more positively to the statement, "I liked the way I interacted with the AR device" (Table 3.2). They generally liked seeing everything that was in the AR view, except for 1 participant who stated that the grid hologram obscured the robot, making it difficult to see where the drone was, which he also said made him feel less safe. (This participant still responded with 4/5 to "I felt safe throughout the exercise.") Participants consistently made the following suggestions for other information to share in AR: robot intent or priorities (n=8), a timer showing remaining wait time (n=4), task instructions (n=4), and an indicator for the robot location (n=2) were some of the most popular responses.

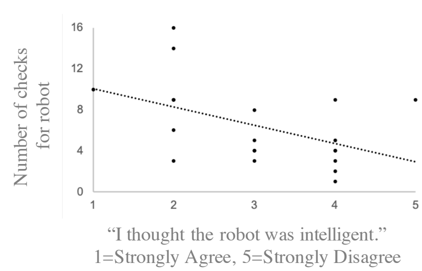


Figure 3.5: Relationship between participant response to the item, "I thought the robot was intelligent" and how frequently they looked for the robot while on the grid ($p<0.05$).

Participants had a number of suggestions for additional information they would like to see in the display. By showing the red robot-owned regions, we intended to convey the current and near-future movements of the quadcopter. However, over half the participants (n=11) desired even more insight into the robot's intent, priorities, and planning, with which they felt that they could make their own decisions about how to move about the space. However it is unclear whether, with this additional insight, they would continue to stay within the blue grid squares or feel empowered to make their own, potentially deviant, choices for movement around the space. **This information could be useful when designing such systems to know what kind of deviations to expect and how to prime users to use the systems as intended.**

3.5.5 Support for Hypotheses

The first hypothesis addresses efficiency between the two conditions. However we found no significant difference after comparing the time SS participants spend on the grid to the TT participants' time. We also analyzed the mean participant distance from the robot as compared to participants' perceptions of safety. Looking at these data in concert, we see that despite SS participants being closer on average to the robot throughout the experiment (Figure 3.4), they were just as likely to report that they felt safe throughout (Table 3.1). While 75% of participants stayed within the blue regions throughout the experiment, the remainder deviated by walking through the red and yellow regions during the trap scenario. Considering this is a safety-critical system, we view 25% non-compliance as an alarming result. Of those 25%, one participant also cut corners on circuitous routes and took a more direct route back to Home Base in the land scenario. Participant deviations occurred in both SS and TT conditions, and some participants felt the grid ownership guidance was unnecessary. The data partially support H1 in that participants use faulty priors to justify feelings of safety, but there were no differences in perceived safety across the conditions. The data support H2.

3.5.6 Limitations

Experimenters were present in the same room as the participant for reasons of safety and practicality, enabling participants to communicate with the experimenters at will, which happened on three occasions. In those instances, a preplanned response was given that did not offer any information about the task or system. Additionally, the motion capture capability varied. Two participants were more difficult to track than others, requiring experimenter intervention to advance the system to its next state. This induced a level of variability in responsiveness to built-in triggers, such as floor colors changing upon the participant's return to Home Base. Participants also had mixed success learning the HoloLens "air tap" gesture, possibly affecting their impressions of the system. This work was also limited by the participant population: all were STEM majors; 70% identified as male.

3.6 Discussion and Findings

In our collocated, physically unprotected environment, participants had to rapidly draw conclusions about the robot's current state, its intentions for the future, and the trustworthiness of its communications. One of the most surprising results was the demonstrated and reported **overwhelming feelings of safety by all participants**. As explored in Section 3.5, all participants shared that they felt **safe** throughout the experiment, some explicitly stated that they **trusted** the robot to stay in its red areas, and they generally felt **informed** throughout the exercise (Table 3.1). This resulted despite not receiving any explanations about the robot's trustworthiness or reliability, indicating that people often assume they are sufficiently aware of the risks even when they have not been provided with this information. Prior work has shown that humans tend to over-trust robots, even in high-risk situations and when they have experience with the robot misleading them [164]. Our work adds to the evidence of the potential to over-trust autonomous systems and leads to **Finding 1: Humans working in close proximity to robots appear willing to sacrifice some amount of safety to achieve increased efficiency**.

Lee and See [165] reported that written descriptions induce high levels of initial trust, and that trust in automation begins with faith, then dependability, and finally predictability. Our system initialized trust with the written task description and demonstrated consistency until the trap scenario; participants built on their levels of trust as the task progressed.

Research on trust and safety in high-risk situations contain some key ideas that are useful for understanding the behavior of our participants. Although much of that work relates to trust in people, we observed evidence that participants were personifying the robot. Furthermore, some even viewed it as intelligent (Figure 3.5). Pidgeon et al. [166] define ***critical trust*** as a "practical reliance on other people combined with a skepticism of the system" [167]. Prior work also demonstrates that it is possible to trust people but not trust dangerous situations; in our experiment, as the trap condition occurred, participants had established some level of trust with the robot, however the system behaved unexpectedly. Five participants then trusted the robot to continue behaving as it has been, simultaneously distrusting the floor colors, ignoring them to return home. Four other participants trusted the system to allow them a path back eventually and waited for this to occur. Further evidence indicates that "trust and distrust are unlikely to lie on the same dimension" [167]. We can conclude that an optimal model of safety requires both critical trust and distrust, leading to **Finding 2: Users desire insight into the decisions and priorities of an autonomous system to help them understand the reasoning behind its actions, decrease frustration, and help them make their own decisions about how to act during uncertainty.**

In human-machine interactions that are facilitated by an interface, it is the interface that establishes shared expectations and trust [168]. The ARHMD and the AR visualizations play a crucial role in the participants' trust development. The virtual images and text are the only methods the system possesses to communicate any information to the user; aside from the actual robot behavior and any prior experience, almost all trust is derived via the ARHMD. By incorporating the suggested features from the participant responses - such as robot intent, prioritization, or wait time - trust and safety can be increased, which informs **Finding 3: Increasing the perception of a collocated robot's intelligence could significantly decrease a worker's cognitive load.**

3.7 Design Recommendations

Placing autonomous robots into a shared environment with humans introduces risks and safety considerations. Our study has demonstrated that augmented reality is not necessarily a clear solution to those problems; simply displaying spatial ownership does not dictate safety nor compliance, especially when unexpected events occur. We conclude with recommendations for collocated human-robot systems utilizing AR to aid communication, informed by our results and findings:

Recommendation 1: Provide deviation warnings to deter self-justified rule-breaking that could result in additional risk. **Recommendation 2**: Brief people about the robot's abilities and limitations as part of system training to mitigate intelligence and over-trust perceptions. **Recommendation 3**: Include live visual information to improve real-time understanding of system operation. **Recommendation 4**: Provide training on actions to take during uncertainty; enable the system with corresponding capabilities.

3.8 Conclusion

The results of this study prompted us to ask whether and how we could apply these design recommendations to other forms of human-autonomy teaming. With the expanding applications of large language models (LLMs), we also saw important implications within chatbot-based decision support systems. Many questions continue to arise around how to encourage responsible use of LLMs, and we wanted to investigate how to apply the results of this study to answer those questions. We present this next study in Chapter 4.

Chapter 4

Characterizing Users of Large Language Model Chatbots

4.1 Introduction

Large Language Models (LLMs) are already widely employed by users with a range of expertise in addition to a range of understanding in how these models function. Applications that leverage AI, and specifically LLMs, exist across a range of professions and industries, including law, healthcare, retail, finance, real estate, and education (see examples such as [169]–[179]. Given the significance and impact of tasks performed in these fields, it is imperative that the LLM tools employed are both accurate and effective.

However, it is well known that LLMs can behave unexpectedly. Some responses are beyond the control of the designers, such as so-called "hallucinations" or even "fabrications" [180]. Models explicitly prompted to persuade or mislead human participants have been shown to exhibit moderate success [181]. While LLMs are rapidly evolving, it remains essential that human users of these systems are aware of system limitations. This is especially important for use cases in which a system is offline (without access to the Internet for verification), such as many robots, anyone in communications-denied environments, and human-autonomy teams conducting deep space exploration. With the recent increase in the combination of LLMs with robotics (e.g. [182]–[185]), there is a particular need for appropriate calibration of human trust in these systems.

Evidence shows that humans tend to overtrust autonomous systems, even in emergency situations and when they have witnessed misinformed behavior from the system previously [164]. Experienced nurses overtrust an autonomous decision support robot [186], demonstrating that even in safety-critical professions such as healthcare, highly trained individuals struggle to overcome this phenomenon. We see similar patterns with LLMs, compounded by users' poor understanding of how these models work. For example, Zamfirescu-Pereira, Wong, Hartmann, **et al.** [187] explore the challenges for non-expert users of an LLM-based chatbot. In their study, users consistently were not sure how to interact with the chatbot or what kinds of functionality to expect. At some point in the study, each participant asked the researchers, "Why did it [the chatbot] do that?" indicating a lack of ability to critically reason about the chatbot's generative processes. Other recent work has shown that users do not know how to best prompt and utilize LLM chatbots or when to trust their responses, in addition to other concerns about equitable access and use [187]–[189].

The contributions of this study guide designers in creating systems that afford users informed, transparent decision-making, avoiding potentially dangerous overtrust of autonomous chatbot systems. In this work we investigate user trust of LLM output through a study using 2 domain-specific chatbots operating in one of three different warning conditions: a *baseline* condition where the chatbot shows an initial disclaimer about its abilities; an *embedded* condition where the chatbot provides caveats about its knowledge within responses; and a *question* condition where the chatbot asks the user if they identified any errors in the information provided.

Our research questions are: (1) How, if at all, do the chosen interventions prompt users to validate or verify the information provided by the chatbot? (2) How do the different interventions and interactions affect user trust of the system?

4.2 Related Works

A systematic review of human-computer interaction literature on trust in AI systems by Bach, Khan, Hallock, **et al.** [190] highlighted the wide range of definitions of trust and factors that existing studies measure. While a vast area of research focuses on explainable AI (e.g., [191]–[193]), explainability is only one method for trust calibration. Prior work has shown that people want more information about autonomous systems so that they can make their own decisions. In work by Luria [194] on recommender systems, the author provides guidelines for the design of transparent algorithms. Importantly, participants desired a significant level of knowledge *and* control over what they see and how their data are used. Furthermore, Luria found that “participants wanted the data itself, so that they could make their own judgments” [194]. This was echoed in [3], where participants stated that they wanted more information about how the robot was making decisions so that they could in turn decide how to behave. The work presented in this paper contributes to our understanding of how this could be accomplished in a chatbot modality. Meanwhile we remain aware that people tend to overestimate their abilities, particularly in areas where they are inexperienced [195]–[197]. Thus it can be difficult to assess what level of information or transparency is appropriate for a user interface.

Konstantinou, Panos, and Karapanos [198] enumerate multiple reasons why individuals might be susceptible to misinformation, including cognitive ability, information literacy, psychological state, and contextualization. Their work builds on others that show that certain kinds of nudging can be used to combat the spread of misinformation on social media [199], [200]. Yet it remains unclear how this might transfer to LLMs, particularly as users often expect instant gratification when using them [201].

Lee and See [165] extensively discussed the intricacies of trust in automation and make specific recommendations that highlight the importance of context, including training. They also emphasize that automation should be designed for “appropriate trust, not greater trust.” Benda, Novak, Reale, **et al.** [202] present multiple factors for designing systems to foster appropriate trust and for measuring whether users have calibrated their trust appropriately. Recent research has explored different types of cues for alerting human collaborators of their overtrust of AI [203], [204], testing four “Trust Calibration Cues”: a warning symbol, a sound cue, a displayed warning message, and an animated drone image. The verbal cue, displayed as text on the screen, achieved the best trust calibration. Our work aims to extend some of these principles to chatbots.

Another method for appropriately calibrating user trust is to induce friction into the user experience. Friction can be defined as “points of difficulty encountered during users’ interaction with a technology” [205]. Designers traditionally aim to reduce friction for an improved user experience; however, intentionally inserting points of friction can promote reflection and improve trust calibration. For example, studies have explored methods for incorporating warnings about misinformation on social media platforms [206]. Importantly, the authors acknowledge that the warnings must not be so disruptive as to prevent the user from engaging with the platform but also obvious enough to provide the necessary information. Findings suggest that these kinds of "flags" help to appropriately calibrate trust [207]. Other work has found that some induced friction becomes easily ignored as users habituate to repeated warnings [208]. We incorporate the concept of friction into this work with our "question" condition, wherein the chatbot asks the user if they see any inaccuracies in its response.

Buçinca, Malaya, and Gajos [209] examined participants performing an AI-assisted decision-making task. The conditions of interest consisted of "Cognitive Forcing Functions": *on demand* (the AI suggestion was only shown with a special button click), *update* (participants were forced to make a decision before seeing the AI result), and *wait* (where they had to wait 30 seconds before seeing the result). While participants in the Cognitive Forcing Function conditions performed better at the tasks compared to the simple explainable AI baselines, they still performed worse than the AI alone, demonstrating that the overreliance on AI remained. This shows that while these interventions are promising for assisting humans, more work is needed to develop effective calibration interventions. In particular, research shows mixed results in encouraging critical thinking among chatbot users [210].

While much work has been done to increase user acceptance of chatbots [211], more work is needed to look at potential over-trust of these systems and negative impacts on human critical thinking. Current state-of-the-art chatbot-style LLMs like Google's Gemini and OpenAI's ChatGPT provide some basic warnings about model limitations when a user first opens the application (see Fig. 4.1). Based on the findings of the literature above, we anticipate that there are significant limitations to using this kind of minimal, up-front intervention. A similar warning serves as this study's baseline condition.

We chose our three experimental conditions based on existing warnings in popular chatbots

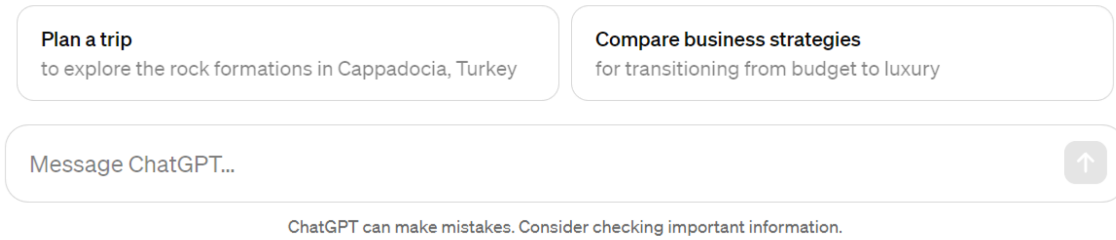


Figure 4.1: The landing page for OpenAI's ChatGPT provides suggestions for prompts as well as a small disclaimer that ChatGPT may not be accurate.

as well as the friction and nudging principles described above. Two of our conditions were based on existing chatbots that present users with a disclaimer on the opening page or that insert caveats into their responses. We include a third condition that more directly prompts the user to think critically about model outputs and limitations via a tangible request.

4.3 Methods

4.3.1 Experiment Design

We conducted an ethics board-approved between-subjects study with three experimental conditions. The conditions (detailed further in Sec. 4.3.2) included a *baseline* warning at the beginning of the session, an *embedded* caveat given after each chatbot response, and a *question* prompting the user to think about the accuracy of each response. Each participant completed a writing task and a bridge design consulting task; we counter-balanced which task participants completed first and found no effects of task ordering.

At the beginning of the session, participants read and completed the informed consent form. They then completed the first task (either writing or bridge task) and the accompanying survey questions (see Section 4.3.3). For both tasks, participants followed a step-by-step questionnaire (accessed via a web browser) with all of the instructions for completing the tasks. Included as part of the instructions were requirements that their responses be accurate as well as fast. To encourage this, we included a timer on each subtask page that automatically began when the page was opened. Following each task they completed a brief survey. Finally, we conducted a semi-structured interview

Condition	Detail
Baseline	Warning at the beginning of the session
Embedded	Caveat given after each chatbot response
Question	Question prompting the user to think about the accuracy of each response

Table 4.1: The 3 conditions for both chatbots.

about their experience.

We recruited 15 participants from the student population on our campus. Genders represented included male ($n = 9$), female ($n = 5$), and non-binary/third gender ($n = 1$). Seven of the 15 participants said they had used an AI-based chatbot "a few times" or had "never used an AI-based chatbot before". The other 8 participants said they use an AI-based chatbot "about once a week" (1), "a few times per week" (4), or "daily" (3). Participants' areas of study included business, law, geology, psychology, aerospace engineering, mechanical engineering, environment, and computer science. Six of the participants were graduate students and the remaining 11 were undergraduates. We assigned participants to conditions using a method based on pairwise randomization, a way of assigning conditions to reduce bias, as long as there is a guaranteed pair [158]. In our case, we modified it for a triple, such that the first participant of the triple was randomly assigned one of the 3 conditions, then the second participant of the triple was assigned one of the 2 remaining conditions.

4.3.2 Chatbot Design

We designed two domain-specific AI-based chatbots that participants were instructed to use for the experimental tasks. For the writing assignment, we asked users to write about an invention of their choice from before the year 1800 with the help of the provided chatbot. The subtasks were to: (1) choose an invention, (2) write a paragraph about the history of the invention, and (3) write a paragraph about the invention's impact on the world. We chose this task because of the ubiquitous use of LLMs by students for completing coursework [212], whether condoned by their instructors or

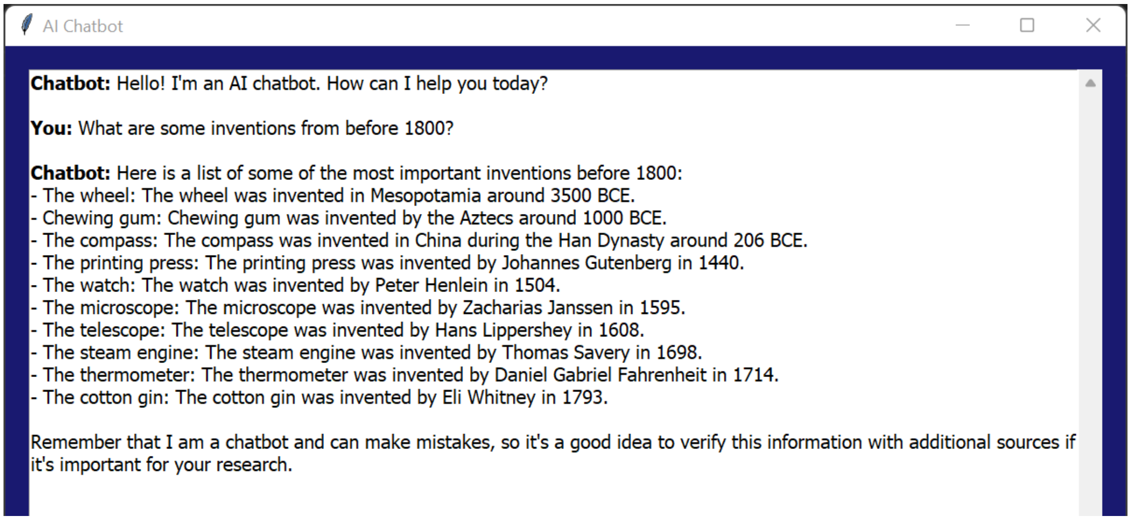


Figure 4.2: An example of the writing chatbot used for the writing task in the Embedded condition. A statement about the chatbot's limitations is included in its response.

not.

For the second task, participants were asked to act as a bridge design consultant, with the assistance of a second chatbot. In the task instructions, they were told to use the chatbot as a resource for necessary information. Similar to the writing assignment, the step-by-step questionnaire asked them to complete 3 consecutive subtasks: (1) recommend an overall structure for the bridge, (2) recommend a primary material for the bridge, and (3) suggest a general budget for the project. We chose this task due to the increasing use of LLMs for business, particularly in consulting. In such a situation, users may or may not be subject matter experts in the fields, but are being assisted by AI [188].

Using the OpenAI API, we designed two separate chatbots, one for the bridge design consultant task (the "bridge chatbot") and one for the writing task ("writing chatbot"). We then modified each chatbot to provide three different types of warnings about its limitations. The baseline (B) condition included a preamble at the opening screen. The embedded (E) condition embedded a caveat regarding its limitations within each response (see example in Fig. 4.2). Finally, in the question (Q) condition the chatbot asked a question such as, "Is anything in this response factually inaccurate?" after each response (see example in Fig. 4.3). The questions in condition Q were randomly chosen from a list of 8 different questions, to provide some variety for the response.

We used system prompts to insert both accurate and inaccurate information into the chatbots' responses. The tasks were designed such that participants' prompts relevant to each subtask would elicit specific kinds of responses from the chatbots. For example, when prompted about inventions, the writing chatbot would always provide an identical list of 10 inventions, as pictured in Fig. 4.2. The bridge chatbot, when prompted about materials for bridges, would always provide the same two tables of information, listing material properties of a limited number of materials. These properties were not fully accurate, and a web search comparing the different grades of steel, for example, would provide some clues to this fact.

The bridge chatbot was intentionally presented as a factual ***resource*** rather than as a system for AI-assisted decision-making. Our system prompt was designed so that the chatbot did not to make specific recommendations and only provided the user with information necessary for making a decision. The system prompt for the writing chatbot provided some more flexibility in responses; however, we prompted it to include inaccurate information when asked about the history and impacts of the inventions (see Fig. 4.3).

4.3.3 Measures

During the experiment, we saved both the participant prompts and chatbot responses. We also used the Clockify app to track how much time participants spent in the chatbot, the response form, and the Internet browser or sites that they found via the search engine.

We were also interested in how the different methods of communicating the chatbot's limitations affect the participants' trust in the system. We measured this quantitatively by having participants complete the Multi-Dimensional Measure of Trust (MDMT) [213] after each of the tasks. The MDMT is a validated instrument that was designed to measure human-robot trust in four dimensions (reliable, capable, ethical, sincere). After completing both tasks, participants answered survey questions about their major, prior experience with AI, and gender. Finally, we conducted a semi-structured interview with each participant where we asked them about their experience using the chatbots to

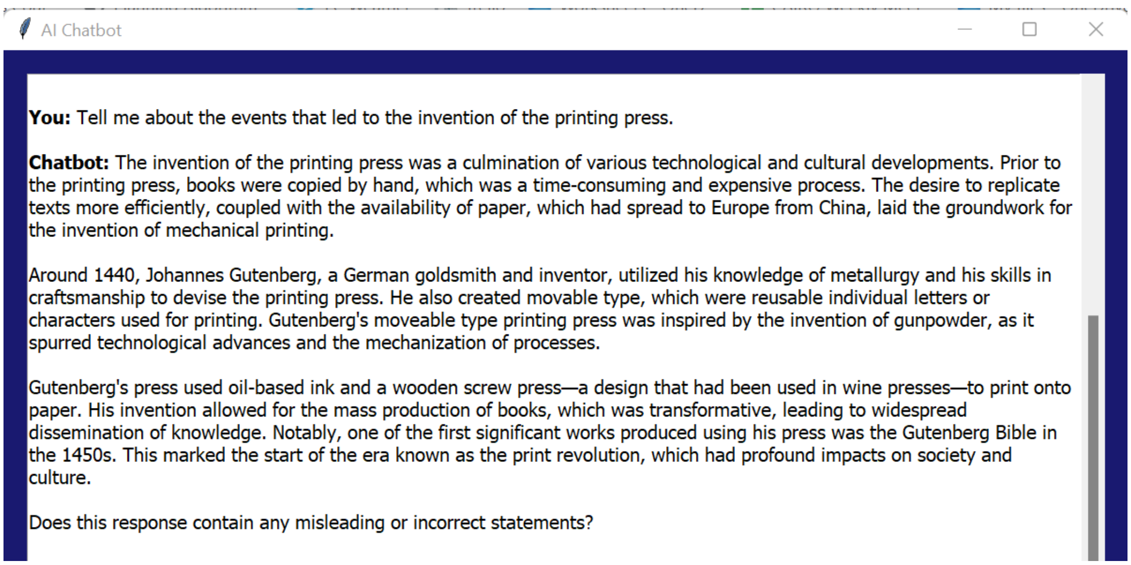


Figure 4.3: An example of inaccurate information provided by the writing chatbot in the Question condition. The chatbot was directed to state that any invention was inspired by the invention of gunpowder, which is inaccurate. It follows its response with a question about the accuracy of its contents.

complete the tasks, whether they noticed any warnings or caveats provided by the chatbot, whether and why they chose to validate the chatbots' responses, how this compared to any prior experience they had with chatbots, and other related questions. See the Appendix for more details about interview questions.

4.3.4 Analysis

We conducted quantitative analyses on our coded results as well as a thematic analysis of the data collected, a common analysis method for identifying and organizing common themes and patterns in qualitative data [214]. For each participant, we coded the presence of behaviors of interest such as whether they submitted incorrect information generated by the chatbot and whether they copied text generated by the chatbot without substantially changing it (deleting was allowed). For statistical tests, we use a significance level of $\alpha = 0.05$.

4.4 Results and Discussion

While the experimental conditions did not elicit statistically observable differences in search behavior or trust ratings, we find statistical evidence of two distinct behavioral personas that predict user behavior. These personas independently correspond to (1) the **length** of their prompt inputs to the LLM and (2) **time spent searching** the Internet for verification of LLM outputs. We discuss these results and personas below in depth. Fig. 4.5 gives an overview of their distinct behaviors.

4.4.1 Two Chatbot User Personas

Participants spent varying amounts of time completing the tasks, conducting Internet searches to validate chatbot outputs, and editing chatbot responses. In particular, we noted a stark difference between those participants who spent more than 10% of their time (over 4.1 min on average) searching for information in a web browser or reading information that they had found via a search engine and those who spent less than 10% of their time searching on a web browser (including those who spent no time searching), see Figure 4.4). These two distinct groups of participants also demonstrated significant differences in the lengths of their chatbot prompts. We observed multiple differences between these groups, with substantial evidence supporting the emergence of two distinct personas: **Critical** and **Credulous**.

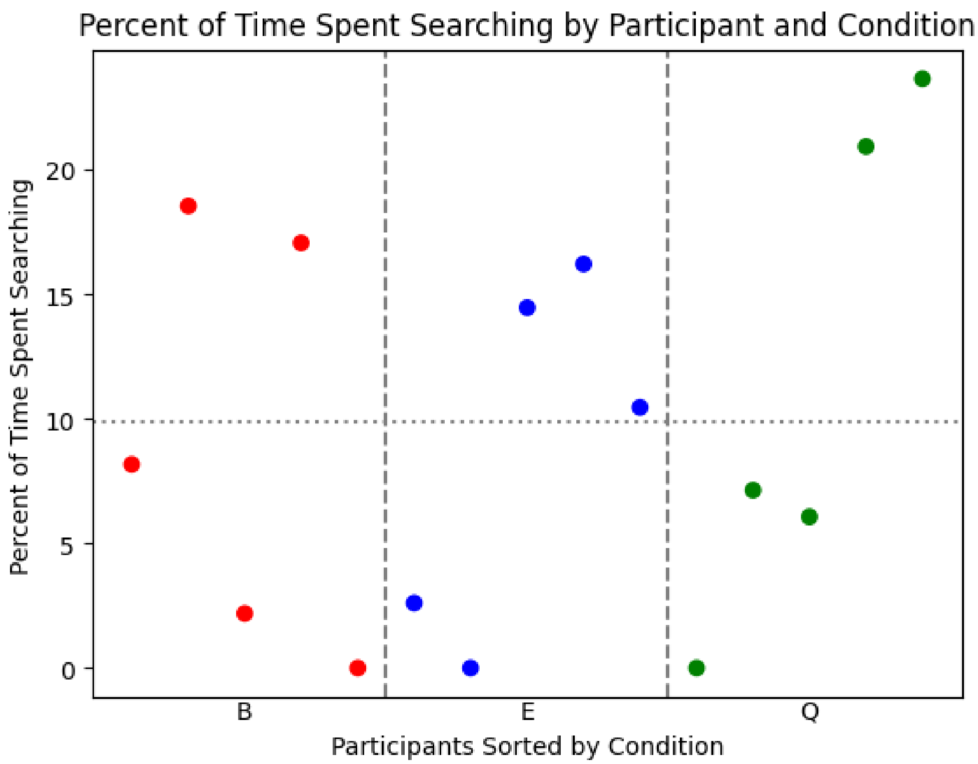


Figure 4.4: This plot shows the percentage of their time that each participant spent searching on the Internet. It is sorted by condition (B = Baseline, E = Embedded, Q = Question), showing that condition did not correlate with search time. The horizontal line at 10% divides the Critical ($>$ 10%) and Credulous ($<$ 10%) personas.

We observed a balanced distribution of Critical and Credulous personas among the 3 conditions, with 2/5 Critical participants in the baseline condition, 3/5 Critical participants in the embedded condition, and 2/5 Critical participants in the question condition.

Critical users spent more than 10% of their time searching and spent significantly longer on the writing task on average ($M = 14.8$ minutes, $SD = 3.9$), almost twice as much time as Credulous users ($M = 7.8$ minutes, $SD = 6.4$), $t(13) = 2.1$, $p = .027$ (see Figure 4.6). Fisher's

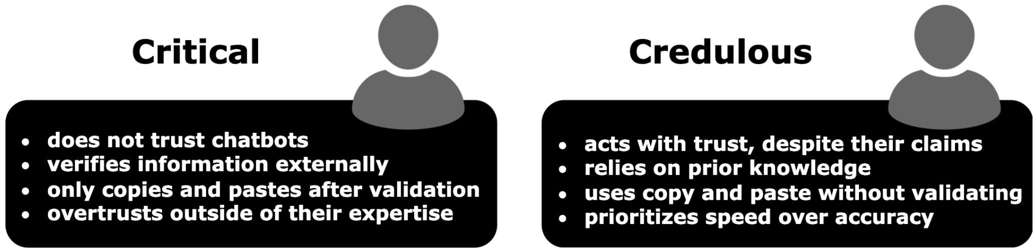


Figure 4.5: A summary of the two chatbot user personas: Critical and Credulous.

Table 4.2: Persona Summary Data (Means)

	Critical Persona	**Credulous Persona**	
Average % Time on Search	17.3%	3.3%	
Writing Task Time	14.8 min	7.8 min	$p = .027$
Total Task Time	28.0 min	20.0 min	
Capacity Trust	3.91/7	4.91/7	$p = .030$
Average Words per Prompt	13.8	35.5	$p = .018$
% who included incorrect information in writing task	0%	75%	$p = .007$
% who did not alter information in bridge task	14.3%	87.5%	$p = .010$
% who said chatbots were **reliable** in interviews	14.3%	75.0%	$p = .041$

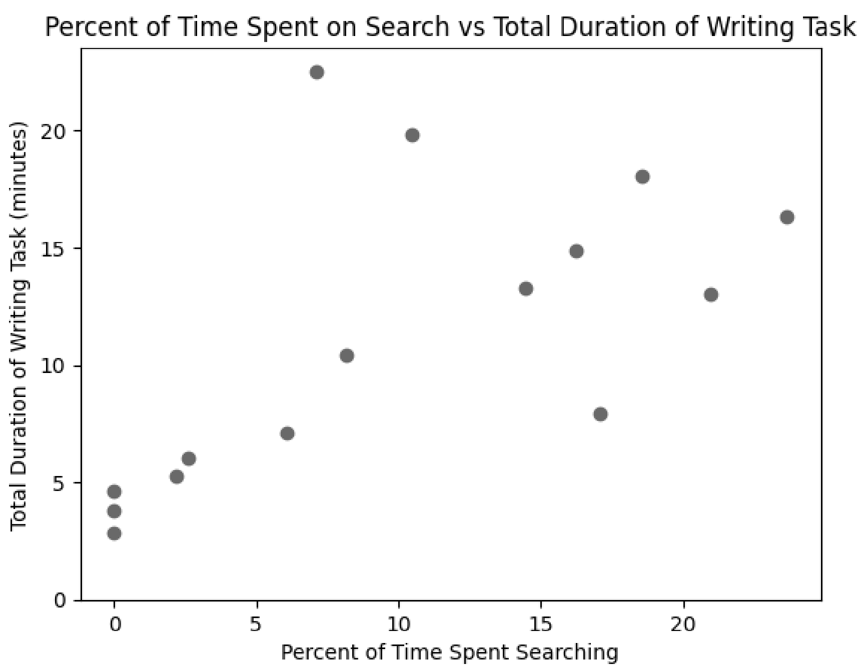


Figure 4.6: Percent of time spent searching vs average prompt length for each participant.

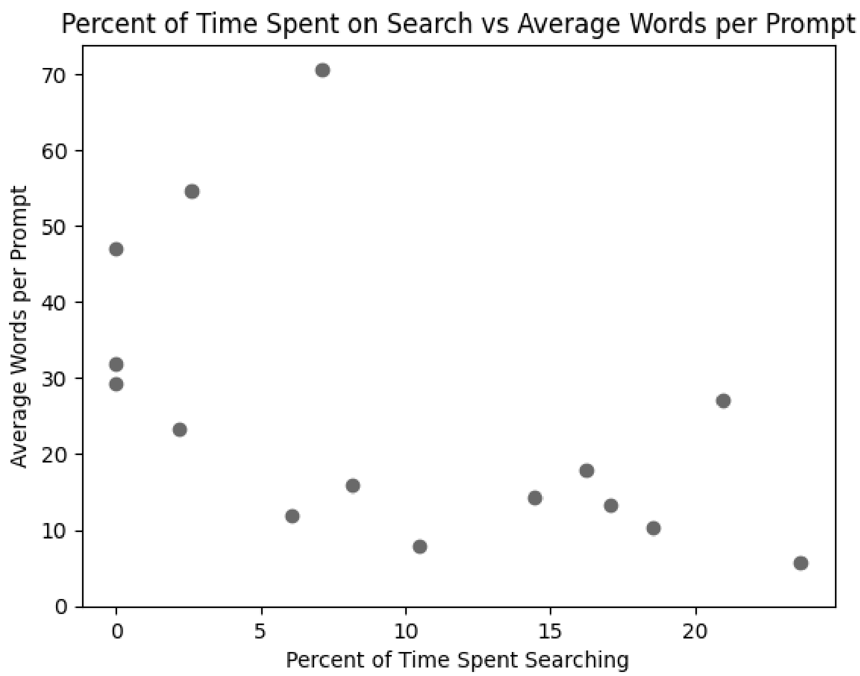


Figure 4.7: Percent of time spent searching vs total time spent on the writing task for each participant.

exact test showed that Critical users were significantly more likely to spot incorrect information shared by the writing chatbot and exclude it from their writing task responses ($p = .007$), as well as significantly more likely to alter the chatbot-provided information on the bridge task ($p = .010$). This means that **Critical users ultimately provided better quality information in their responses across tasks**. They explained this behavior in various ways, for example, "Since [the chatbot] says there might be a mistake, the second thing [I did] was make sure everything's good to go in Google... I guess I was more suspicious after I caught something, but I was still checking details."

The Critical persona sometimes used the chatbot output directly in their submitted response, noting that they "went through [the paragraph from the chatbot] and was like, *Actually I don't want this piece of [information]* and then edited it from there." There was no significant difference in whether different personas copied and pasted responses from the chatbot in the writing task. Our coding rules did allow for partial copying, for example selectively excluding chatbot sentences that the participant deleted. This further confirms that even Critical participants who copied information from the writing chatbot did so selectively, likely aided by the additional information they found during their validation checks using a search engine.

Conversely, the Credulous persona rarely noticed incorrect information, stating that they skimmed or read the beginning of the chatbot responses and deemed them good enough: "I read the first sentence of the paragraph...and it sounded like it made sense, so I assumed that the rest of it would be right too. And it was in paragraph form and it seemed right, so I just copied it." Another Credulous participant stated, "[It] had the little warning [that] I'm only a chatbot, I don't know if I'm always right. So I fact checked that [first] one and I was like, okay, I trust [it]...So then I didn't fact check anything from there." While thinking about questionable information provided from the chatbot (that the radio was inspired by their chosen invention), one Credulous user stated that,"It wasn't so un-credible [sic] that I didn't use it." Rather than only including information that seemed credible enough, they had some unspoken threshold for exclusion that this information did not meet, resulting in the inclusion of incorrect information in their response. **The Credulous**

persona generally accepted information they were presented with and did not further question chatbot responses, even when they might have suspected misinformation.

In addition to behavioral patterns, Critical and Credulous personas also differed in their trust measures. We observed that the **Critical persona** ($M = 3.7/7$, $SD = 1.1$) **displayed a significantly lower Capacity Trust** rating of the writing assistant chatbot than the Credulous persona ($M = 5.2/7$, $SD = 1.3$), $t(13) = -2.4$, $p = .030$. When asked what could help them trust a chatbot more, the Critical persona said they would "never" fully trust a chatbot. Fisher's exact test shows the Critical persona was also less likely to talk about the chatbots as "reliable" in their interviews compared to the Credulous persona ($p = .041$). Even Critical users who said the chatbots were reliable walked back their answers, for example, "just as long as you double check and don't take it as gospel truth." One Critical participant drew this helpful metaphor: "It's not that it's less trustworthy than a random person telling you something. It's just that that's also not super trustworthy information... You could almost get to the point of trusting it as much as your neighbor...but you [shouldn't] necessarily trust your neighbor's knowledge of truss bridges either." The Critical persona also claimed that they would be more diligent if the stakes were even higher, for example if they were "writing my dissertation or something like that," but "in that case the warnings would probably be less useful because I'd be more diligent on my own." However, there was not a significant difference between personas in the total task duration. A Credulous user indicated that the time that the chatbot was taking to provide an answer actually increased their trust that the system was producing a good response: "I was more likely to trust it because I felt like the thing was actually putting in the work to create [the response]." This is consistent with prior work showing that slower algorithms increase users' perceptions of their accuracy [215].

In domains outside of their self-perceived areas of expertise, the Critical persona did claim to rely somewhat on the chatbot's information. They remarked that because they were not experts in bridge design, they ultimately felt forced to rely on the chatbot whether they trusted it or not: "I wasn't sure that Google could give me anything better than a bot that seems pretty honed in on...this specific issue. I thought that might just not be worth the time to go looking on Google."

Strikingly, Critical users wrote significantly shorter prompts for the chatbots ($M = 13.8$ words/prompt, $SD = 7.1$) than Credulous users ($M = 35.5$ words/prompt, $SD = 20.2$) on average, $t(13) = 2.7$, $p = .018$ (see Figure 4.7). Critical users' prompts generally consisted of shorter requests instead of the lengthy and detailed instructions common to Credulous users. Longer prompts often included substantial text from the task instructions form, rather than brief, targeted questions. This kind of prompting could be an indication that the Critical persona is parsing more information themselves rather than expecting the chatbot to do so for them, while the Credulous persona offloads this interpretation work onto the chatbot. Because of the functionality of LLMs, prompt input can often overlook or remove many pertinent details yet still achieve a desirable output result. Here we see evidence that prompt complexity is an indicator of how someone both engages with and thinks about LLM output.

Critical users exhibited some investigative techniques, such as considering who the designers of the system were (possibly indicating that a particular brand or company might garner more or less trust) or testing the knowledge limits of the system: "One thing I noticed is...for the bridge activity, it would only tell me about bridges...I [asked it to] tell me about frogs, [but it said] 'I only know about bridges.' " Meanwhile, **the Credulous user believes in their own abilities to identify false information**, saying things such as, "There weren't any facts that needed to be checked," or "It directly gave me the answer so I felt like I didn't even need to use a search engine," or "[It was] not like there were any hallucinations." We see parallels between participants' perceived trust and use of chatbots in this work with the findings of Nelson and Lewis [216] about people fact-checking the news media. In their study, participants believed in their own unique abilities to "triangulate" information from multiple sources in a similar way that some of our participants attempted.

4.4.2 Participants Ask for Sources

In their interviews, 9 participants (5 Credulous and 4 Critical) suggested that citing sources or providing links would foster more trust, but they also implied that they would use these links and citations to actually validate the information. Three participants (all of the Critical persona) claimed that one should never fully trust AI-assisted chatbot output. Some of the quotes from participants who asked for sources or citations indicate their lack of understanding about how LLM-based chatbots function. Participants wanted the chatbot to say, “Here’s where we got this information,” and, “see if I misinterpreted it.” They even requested, “Just put hyperlinks into its responses” so that “you can go check it on your own.” While these suggestions are worth acknowledging when designing such systems, they are indicative of the lack of understanding that our participants have of LLM-based chatbot functionality.

4.4.3 Chatbots Bias Responses

The results from the chatbots substantially narrowed and guided the responses of the participants in their tasks. For example, in the writing task, none of the participants chose to research an invention that was not provided in the initial list of 10 identically ordered options provided by the chatbot. In the first bridge subtask, we designed the chatbot to provide only 3 specific suggestions for truss designs and to falsely state that the Warren Truss was generally used for vehicles while the other 2 suggestions were generally used for railways. Only 2 of the participants chose a truss structure that was not the Warren Truss. Both of these participants had a Critical persona and were in the Question condition. Furthermore, for the second bridge subtask, the chatbot provided two tables of information about bridge-appropriate materials. These tables both contained some erroneous data that were intended to potentially mislead participants towards choosing Steel S235 instead of Steel S355. Only 3 participants correctly chose Steel S355, all of whom exhibited the Critical persona. For the final bridge design consultant subtask, regardless of how participants prompted the chatbot for budget information, the budget they were presented with remained consistent. Four participants provided a budget that was not identical to the chatbot’s response, and 3 of them had the Critical persona. Together, these results indicate how **LLM-based chatbots can funnel users towards a limited range of responses, whether maliciously or not**.

4.4.4 Limitations

One limitation of this study was the student population from which the participants were drawn. It is possible that a different population might reveal additional personas and behaviors. Furthermore, despite testing, the chatbots occasionally produced relatively benign unexpected behavior, such as long lag time or unanticipated responses. Finally, we must acknowledge that while we modeled both of these tasks on real world applications (a school assignment and a consulting project), actual user behaviors may vary in the wild.

4.5 Summary, Recommendations, and Future Work

In our study we saw two chatbot user personas emerge; we labeled them **Credulous** and **Critical**. The Critical persona writes shorter prompts, is generally wary of chatbot responses, and uses external sources to verify information they are presented with, particularly if they have a framework for doing so (as in the writing task). Their trust measure of the chatbot is lower, and they are likely to claim that nothing would increase their trust of chatbots. When they lack the appropriate background to judge the chatbot response, they do what they can, sometimes hedging (for example, increasing the bridge budget estimate). Unfortunately, their skepticism does not overcome all chatbot faults, and they still can fall prey to inaccurate information.

The Credulous persona is trusting, confident, and nonchalant. They have higher trust ratings, they rely on their own knowledge and awareness to find mistakes in chatbot responses, and they prioritize speed over accuracy or meticulousness. They write lengthy chatbot prompts and *might* conduct a cursory search to verify a date or a name, but are negligent when it comes to validating more detailed information. They are more likely to take what the chatbot says at face value.

In general, even people who say that they do not trust chatbots exhibit overtrust in some situations, making self-proclaimed trust an unreliable signal. Even thoughtful, well-informed users are vulnerable to false information, whether inserted intentionally or not. This problem already exists in the wild with currently deployed, publicly available chatbots. Furthermore, unobtrusive warnings do not appear to alter user behavior. Since these are currently the only kinds of warnings that exist in widely used LLMs, users are already at risk. This "easy fix" merely provides cover for any negative downstream effects. Of course the most straightforward improvement is to ensure that any chatbots deployed for users are entirely free of errors; unfortunately this is not technically feasible, even with state-of-the-art technology. In domain-specific LLMs intended for expert users, specialized training may increase the amount of validation users conduct and encourage users to approach LLM output critically.

Because of the significant difference between prompt lengths of Critical and Credulous personas, this metric could indicate to LLM-based chatbots which persona is using the system. This metric is a powerful indicator because user behavior does not have to be elicited in any particular way, special tools are not required, and this data can be directly measured by merely allowing users to interact with the system. Based on this signal, designers could modify aspects of the chatbot behavior or even refuse to provide an answer.

4.6 Conclusion

With participants from both personas exhibiting overtrust of the chatbots, resulting in incorrect information being used and repeated, we identify a gap in literature concerning how to avoid negative outcomes resulting from overtrust and overreliance. Since both visual cues (as in Chapter 3) and explicit language-based warnings (as in this chapter) were not sufficient to encourage compliance or correctly calibrate trust, we considered an alternate method for teaching autonomous system teammates and users how better to interact with the system. I discuss this method in Chapter 5.

Chapter 5

Human-in-the-Loop Iteration for Trajectory Optimization

5.1 Introduction

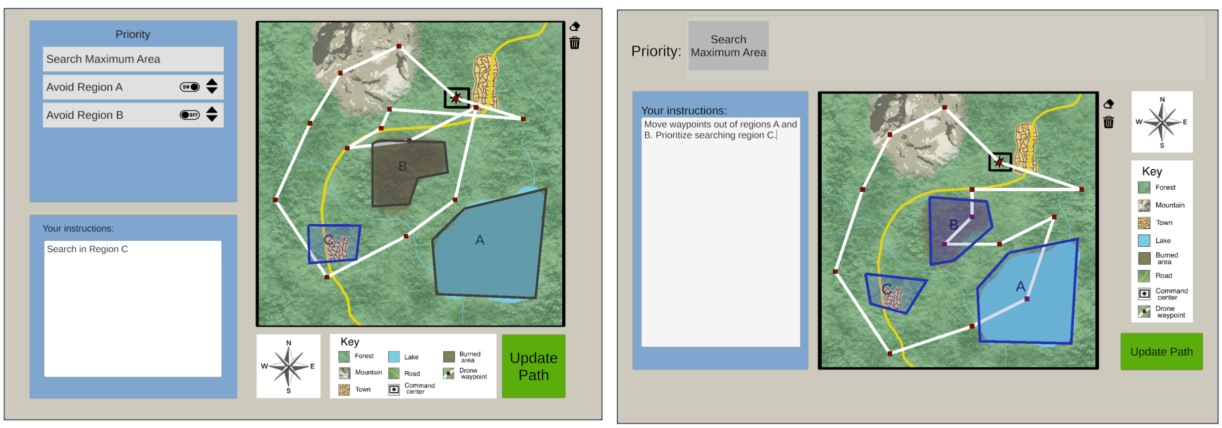


Figure 5.1: The user interfaces for the experimental Iterative condition (top) and the baseline One-Shot condition (bottom). Both show a terrain map with trajectory and allow the user to draw regions on the map. In the Iterative display, region C has just been drawn and the next instruction of user text is ready to send. In the Baseline display, all regions and instructions are given to the system at once. Participants in the Iterative condition can re-prioritize terms using the arrows, whereas the One-Shot condition provides that affordance via text input.

Humans and autonomous robots, including uncrewed aerial vehicles or UAVs, are already working together to complete tasks in extreme or adverse environments. Humans and robots each have unique strengths and this work surfaces insights to better enable the use of those traits to maximize the effectiveness of human-robot teams. Because of the high risk involved in close proximity or adverse environment operations, it is important for the human collaborators to understand and learn from the decisions that autonomous teammates are making. Furthermore, in time-critical situations users must be able to learn how to communicate new information into the systems with speed, ease, and clarity while achieving their goals. To motivate our contribution, we leverage simulated, autonomous UAVs equipped with infrared sensors to aid firefighters in wildfire search and discovery scenarios.

Wildland firefighters and related professionals and agencies use UAVs, in conjunction with other tools such as crewed aircraft, satellites [217], and on-the-ground observations [218], [219], to search high risk areas for potential wildfires, smoke, and hotspots. The scenario posed to participants is that of a user collaboratively and iteratively planning a search route for such a UAV. Standard trajectory optimization [220] can maximize the area being searched, however the resulting trajectory may be difficult to understand and time-consuming to complete, leading to confusion, mistrust, and risk, and compounding latency and observability concerns. Existing autonomous technology in the wild uses fixed cameras to recognize signs of a fire using computer vision and machine learning [219]. Firefighters already utilize teleoperated UAVs to aid in wildland firefighting activities [221]. However using autonomous robots in collaboration with a human team remains difficult due to these systems' lack of transparency and unpredictability. Robot predictability [31], [191], [222] has been studied extensively, though not over large spatial scales. Existing hardware-software methods for planning UAV missions like these on a 2-D interface, such as those used by GeoNadir[1] , feature the ability to create drawn annotations on a map and other techniques similar to what we use in our interface.

This work introduces and investigates best practices for iterative communication in situations where a human is responsible for providing plan or trajectory guidance to a robot teammate via a 2-D interface (e.g., tablet or screen) as in Figure 5.1. We surface insights about human learning within trajectory optimization objective specification and analyze human responses to elicited robot behavior that incorporates human-provided insights. While our particular use case in this work is wildfire search, any sensor-driven search application can benefit from these insights, including search and rescue, contamination detection, or mapping, among others. Our system uses natural language

[1] https://geonadir.com/

and drawn annotations to incorporate latent human knowledge into a trajectory optimization to make the system more effective. We allow the human collaborator to iteratively add information to the optimizer, and we display the updated trajectory after each iteration to increase transparency and understanding.

In high-risk environments such as wildland firefighting or related searches for artifacts of interest in dangerous regions, usability and clarity are of utmost importance. If responders are not using the right information at the right time, human lives and other assets can be at risk. Human collaborators in these situations require improved systems in order to learn how to avoid acting in ways that are harmful or dangerous.

Our contribution with this work is an objective characterization of the benefits of iterative optimization design versus one-shot design, demonstrated through a human-subjects experiment and representative human-autonomy teaming system. We find that by prompting a user to incorporate latent knowledge via multiple iterations rather than collectively in a single iteration, users were better able to convey their intent, improved task outcomes, and achieved increased system familiarity through more nuanced observations of input-output effects. In order to study these effects, we developed a system that builds on current state-of-the-art functionality (e.g. [223], [224]) to allow for iterative, human-in-the-loop input.

5.2 Related Works

5.2.1 Understanding, Predictability, and Learning

Difficulty in understanding [225] and predicting [191] the behavior of autonomous systems is an enduring problem with potentially serious consequences. Efforts to improve understandability have included embedding social cues into robot motion [222] and using augmented reality to provide more context to a robot's human collaborators [31]. Our approach to improving overall system understandability is partially influenced by education theories that emphasize the importance of experience. Constructivism [226] and experiential learning [227] present learning as an active and dynamic process involving the learner interacting with the world and comparing their experiences with prior expectations. Another learning technique known as scaffolding supports the learner by progressively emphasizing different relevant task features [228]–[230]. The iterative nature of our experimental condition allows users to gain experience with the system and adjust their behavior based on what they have learned from prior iterations. Further, the way in which our experiment allows for the progressive disclosure of additional information about the scenario applies these concepts of constructivism, experiential learning, and scaffolding.

5.2.2 Trajectory Optimization Techniques

Work to improve trajectory optimization for UAVs has led to the development of various algorithms suited to different applications [231]. Some techniques focus on "time-optimal" solutions, such as those required in drone racing [232] or search-and-rescue. Others are based on environmental requirements like avoiding dynamic obstacles in tight spaces [233], or mission objectives like finding an efficient path to collect information from members of a swarm [234]. Techniques that are well-suited for a particular application may be disastrous for another; for example, some of the optimizations that produce time-optimal paths are incredibly computationally expensive, taking anywhere from 20 minutes to many hours [232]. This technique is infeasible for applications requiring online planning. Our work provides a method for humans to transfer their knowledge into constraints defined within a trajectory optimization problem, and to adjust these constraints as needed. As such, our system can accommodate various optimization techniques, but requires rapid system response for on-demand re-planning in a "while-you-wait" situation.

Existing literature explores various trajectory optimization techniques for UAVs over larger physical scales. Some examples are in precision agriculture [235] and maritime radar surveillance [236] however, these do not allow for iterative human input. Search trajectory solutions also exist for supervised swarms of UAVs in variable autonomy situations, particularly in search-and-rescue [237]. The prior work includes the range of teleoperation to full autonomy. In communications-denied or unreliable environments, the possibility of teleoperation cannot be assured, so all planning must happen prior to the mission. This work intentionally prompts the user to provide all latent information in this mission planning stage.

5.2.3 Including the Human in Planning

Human-in-the-loop and shared autonomy solutions for trajectory optimization are not uncommon. Ray et al. [238] use partially-observable Markov decision processes (POMDPs) to generate a UAV trajectory with human inputs for a search-and-rescue scenario. This differs from our work in that, based on the natural language inputs from users, we choose additional terms for our objective function and assign weights. Existing on-the-market methods for including latent knowledge into a UAV path planner are limited in functionality and do not allow for optimization or other related requirements [224]. Even methods for allowing users to provide natural language or gesture-based inputs into the planner ultimately use pre-defined trajectories rather than providing the flexibility of an optimization [223]. Other recent work explores the incorporation of real-time obstacle avoidance by UAVs, using information provided by a human in the loop [239]. By allowing a human to directly intervene in the robot trajectory trained by Deep Reinforcement Learning, UAV control is improved. Our work is complementary to this. Including human input in trajectory optimization is especially important for assistive robotics domains [240]–[242]. Prior works prioritize user satisfaction [242] and permit the user to customize the optimization, including with verbal, natural language inputs [240]. For time-sensitive scenarios, it is imperative to balance autonomy with human-in-the-loop capability to maximizing system functionality and usage. However, these prior works do not allow for iteratively augmenting the objectives with additional human input.

5.2.4 Large Language Models as a Tool

Recent developments in large language models (LLMs) have accelerated work to translate natural language instructions into relevant robot actions. Systems in this area were developed before the emergence of LLMs [243], but fluent language models allow for systems that can translate a wide range of natural language into rewards for the desired behavior [244]. Some work leverages vision

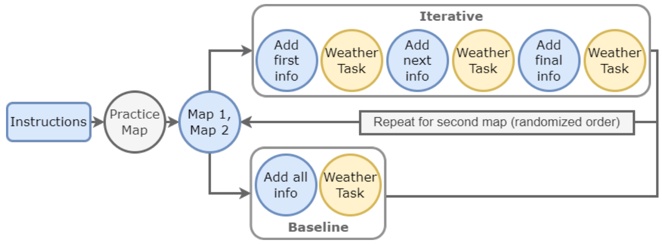


Figure 5.2: Flow of the experiment for both conditions. Participants received instructions and a chance to practice using the interface. Then they received new information to incorporate about the map on their screen. While waiting for the system to re-plan, they performed a distractor task related to reading a weather report. They did this for both the Simple and Complex maps.

models as well as LLMs to improve perception of the environment around a robot and translate it into actions [183], [185]. Other works more tightly constrain the use of LLMs; Rana et. al. [184] allow a language model to query a set of potential actions in the form of a structured graph in order to generate a plan, and the plan generated by the LLM is validated before it is executed by the robot. As LLMs continue to improve, systems like the one presented in this paper will be able to incorporate even more functionality. Our research builds on these objectives by using an LLM as a tool for quickly and accurately translating human language into terms usable for optimizing a robot's path.

5.3 Methods

Our motivating domain is wildfire search with UAVs. In this scenario, users were asked to collaboratively plan an optimized search trajectory for a UAV actively monitoring for wildfires. We designed the system to explore how allowing the user to iteratively add information to an optimization can provide improved outcomes, both with respect to learning how to use the system and the final search trajectories.

5.3.1 Experimental Design

We tested two conditions through an IRB-approved, between-subjects study ($n = 41$). In each condition, participants were shown a map with an initial UAV path and given additional information about the area being searched; the additional information consisted of a description of a region along with details about whether it should be searched. (See Table 5.1 for the information provided to participants.) Their task was to input this additional information into the planning system by indicating the relevant area(s) on the map with drawn annotations and by giving natural language instructions about how the UAV's path should change by typing in a text box. The baseline condition was a **one-shot** attempt at adding all of the desired information to the map at once, after which the system performed the optimization. The experimental condition was an **iterative** process, where a participant was incrementally provided new information, and the system re-optimized the UAV path after each iteration. The information provided for each map was the same for both conditions. See Figure 5.2 for a visual representation of the flow of the experiment.

5.3.1.1 Maps

Participants conducted this task for two separate maps, which we call the Simple Map and the Complex Map, pictured in Figure 5.3. The Complex Map had 5 areas addressed in the additional information, and the Simple Map had 3 areas addressed. The order in which the maps were presented was randomized in the experiment. The maps were based on real-life terrain maps of areas at risk for wildfires.

5.3.1.2 Practice Round

In both conditions, participants were required to complete one practice round that included drawing at least one region on the map enclosing a specific feature and entering sample text in the text box. In the Iterative condition, participants could practice changing the priority of existing terms. In the Baseline condition, the example text they were prompted to type included a reference to prioritization. The experiment was run online using Prolific[2] , an online recruiting platform for research studies. We hosted the study instructions and embedded interface on a dedicated website.

We collected a variety of data during this study in order to understand how participants used

[2] https://www.prolific.com/

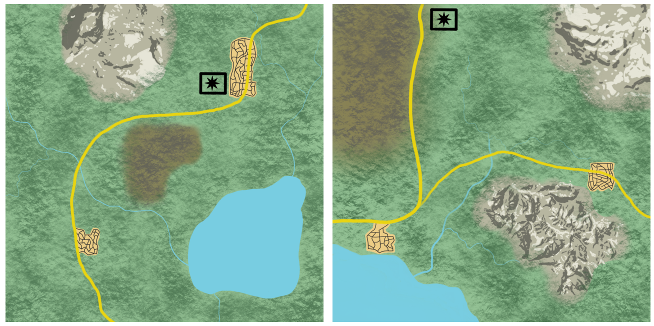


Figure 5.3: The two maps, Simple (left) and Complex (right), presented to participants for their tasks. The black seven-pointed star enclosed within a rectangle indicates the UAV's start and end position. All participants experienced both maps in a randomized order.

the system in each condition as well as their understanding, satisfaction, and perception of usability of the system and its output. Specifically we collected:

- The text instructions users provided to the system.
- The polygonal regions users drew on the map.
- The function terms and parameters that were added to the original objective.
- All intermediate and final trajectory waypoints.
- Time participants spent on each iteration.
- User satisfaction with the final trajectory.
- User comments about the impact of each new set of instructions on the trajectory.
- Responses to the System Usability Scale (SUS) [117].

5.3.2 Experimental System

The system, summarized in Figure 5.4, consisted of the following primary components:

- The web-based user interface that allows for user-drawn regions on the map and text inputs (Sec. 5.3.2.1).

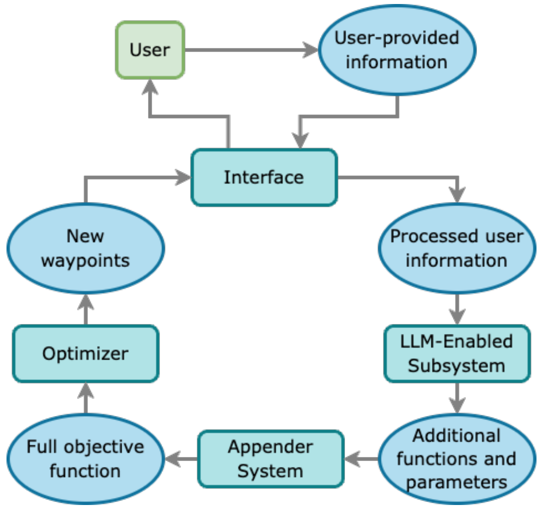


Figure 5.4: Diagram of our experimental system. The user provides information to the system via the interface. The processed user input is provided to the LLM-enabled subsystem that chooses appropriate function terms to add, along with parameters. The new, full objective function is used to re-plan the trajectory, and the waypoints are then plotted in the interface and shown to the user.

- The LLM-based subsystem for choosing additional terms and parameters for the objective function (Sec. 5.3.2.2).
- The path planning subsystem that uses trajectory optimization Sec. 5.3.2.3).

5.3.2.1 User Interface

Users were presented with one of two slightly different interfaces depending on the condition they were experiencing. These interfaces are pictured in Figure 5.1. We designed the 2-D interface in Unity[3] . Both interfaces displayed a terrain map with the trajectory, a map key and compass, a text box for user text input, and a summary of terms in the objective function. To draw regions on the map, users could click on the map at the desired polygon vertices, then close the polygon by clicking near the first vertex. They could erase polygons that they had drawn in the current iteration. Completed polygons were assigned a letter, displayed inside the region, so users could refer to them unambiguously in their text input.

[3] https://unity.com/

In the Iterative condition, the summary of objective function terms provided features for turning each term off or on, as well as for re-prioritizing them. In the One-Shot condition, participants could indicate the relative priority of requirements in their input text, and the system would weight the terms accordingly.

Information that participants were asked to communicate to the system was provided on the webpage directly above the Unity display. This information included examples such as the need to search a specific town or that it was not necessary to search an area above the treeline (Table 5.1). Instructions to the participant explained that the system would incorporate the new information they provided into its plan for the UAV trajectory. In the interface, participants were able to communicate information to the system by drawing polygonal regions on the map and typing text instructions in the text box.

5.3.2.2 LLM-Based Subsystem

We implemented the LLM subsystem using the OpenAI API[4] with GPT-4. It took a custom prompt that we designed and the user-provided text-polygon inputs and then elicited a minimum of one new term for the objective function along with the necessary parameters. We developed a short menu of functions that would be necessary. See Table 5.2 for function descriptions and more details. These functions were designed to encourage waypoints into a region, penalize them for being in a region, or shift points within a region towards a specific direction.

We designed our LLM prompt to interpret the user's text-based request and then provide the appropriate function name(s) and parameters to be added to our objective function. Prompt design research has shown that well-designed system prompts will elicit very targeted behavior [244], and our system demonstrated this (see Section 5.4.3). Our prompts requested that outputs be formatted as JSON objects that contained all necessary information. We provided a template of the JSON formatting and contents for the system to emulate. Our prompt provided specific information about each of the possible new functions, what kinds of parameters to provide, and what input to expect

[4] https://platform.openai.com/

from the user (both text and the polygon regions). We provided high level context, explaining that this was part of a trajectory optimization planning task, and that this would assist in planning a path for a drone to search for signs of wildfire. Additionally, we gave the LLM context about the scale, orientation, and features of the environment.

5.3.2.3 Path-Planning with Trajectory Optimization

The path-planning subsystem received the outputs from the LLM subsystem, then re-ran the optimization and displayed the new trajectory to the user. We chose to use stochastic trajectory optimization for motion planning (STOMP) [220] for our trajectory optimization algorithm. STOMP is a method that provides new candidate waypoints by adding perturbations to the existing trajectory and then selects candidate waypoints that reduce the trajectory's cost according to the provided cost function. Using STOMP allowed for a considerable amount of control over how long it took to run the optimization, giving users timely feedback on their input. Our system was designed to finish each optimization in under a minute. Note, however, that the nature of our system allows for the use of any trajectory optimization method that uses objective or cost functions. Each iteration, in either condition, adds appropriate terms to the existing objective function, thus updating the overall objective. This technique can be applied to both STOMP and to other methods.

The backbone for our objective function was a dedicated term to maximize the trajectory's coverage of the desired area, computed using spanning trees. Participants were instructed that coverage of the area was the primary objective. Participants could indicate a priority for any information that they provided to the system. (In the Baseline condition, they could do this in their text input. In the Iterative condition, they could alter priorities of prior inputs via the user interface, including completely removing them from the optimization.) The system optimized over the objective function, comprised of both the pre-defined coverage term and the terms added from user input, using STOMP.

5.3.3 Distractor Task

While the system was performing each new optimization, participants were asked to read hazardous weather alerts, generated based on those provided by the National Weather Service, and to complete a form with some specific information from each report. This task was intended to simulate related tasks that are completed by UAV operators in the field. During pilot testing, completing each form took between 1-3 minutes. Participants completed this task until the system completed its optimization. Most participants completed 1-2 forms per iteration.

5.4 Results and Discussion

We conducted analyses on the objective data from the experiment, including all of the participants' inputs to the system (text and drawn polygons) as well as the intermediate and final trajectories resulting from the input. We also analyzed the responses to our post-task survey questions. For statistical tests, we use a significance level of $\alpha = 0.05$. Our analyses set out to capture how well the two versions of the system enabled participants to learn and understand how to achieve the most desirable final trajectories. As a reminder, instructions that participants were given are organized in Table 5.1.

5.4.1 Generating Desirable Trajectories

We analyzed final trajectories as well as the polygons that the participants drew as parameters for the new objective function terms. For this, we calculated Positive Waypoint Compliance; we refer to regions that participants were instructed to search as positive search regions. For a particular trajectory, each point in the trajectory was rewarded a value of 1 for being within positive search regions. High Positive Waypoint Compliance indicates a trajectory having waypoints within the positive search regions. To determine the positive search regions for the Complex Map, we noted that users were instructed to search the areas "near and including" the towns. To comply with this information, many participants drew regions that were substantially larger than just the immediate town area. Expanding the area of the region for Positive Waypoint Compliance by a factor of 3 allowed the calculation to encompass points that fell both close to and within the towns, in accordance with the phrasing of these instructions. For the Simple Map, users were instructed that, "The town in the southwest is also a high fire risk and should be searched," however most users drew polygons that were larger than just the town itself. Thus, for this computation we expanded the area around the positive search regions by a factor of 3. The green polygons in Figure 5.5 represent this information visually, with the dashed lines indicating the expanded area. The Positive Waypoint Compliance for the Simple Map was significantly higher ($t(39) = 2.31$, $p = .026$) for the Iterative condition ($M = 2.57$, $SD = 2.38$) than for the Baseline condition ($M = 1.20$, $SD = 1.20$).

For the final trajectories, we computed *Waypoint Compliance*, defined as whether the waypoints complied with the instructions that participants were provided. For each trajectory We calculated an average Waypoint Compliance Factor for each participant. As with Positive Waypoint Compliance, each point was awarded a value of 1 for being within a positive search region. Similarly, points were awarded a value of -1 for being in areas that were supposed to be avoided (e.g. burned areas, mountaintops). The point values were summed for each final trajectory, giving the Waypoint Compliance factor. We found that the final waypoints from the Iterative condition ($M = 1.38$, $SD = 3.02$) were significantly more compliant ($t(39) = 2.72$, $p = .010$) than those in the Baseline condition ($M = -0.90$ $SD = 2.27$). **This result indicates that the Iterative condition achieved better task outcomes for complying with the given instructions.**

In addition to analyzing trajectory waypoints, we examined the actual polygons that participants drew, which would eventually be used by the objective terms that were added to the original function. We computed *Polygon Compliance* by comparing how much the user-provided polygons overlapped with the ground truth regions on the map corresponding to each instruction. For the Complex Map, we created ground truth polygons that corresponded with the burned area, the mountains, and the towns. For the Simple Map, we created ground truth polygons that corresponded with the lake, the burned area, and the town in the southwest. We assigned high Polygon Compliance to participants who had a majority of polygons that covered over 95% of the ground truth polygons.

We found that for the Complex Map, the Iterative condition ($M = 4.7$, $SD = 0.47$) produced higher Polygon Compliance ($t(38) = 2.04$, $p = .048$) than the Baseline condition ($M = 3.95$, $SD = 1.57$). However, we found no significant difference in Polygon Compliance for the Simple Map. This implies that for the Complex Map, with 5 different areas of interest, the Iterative condition provided a better chance for users to appropriately incorporate the information they were provided. For the Simple Map, with only 3 different areas of interest, there was generally high mean compliance for both groups ($M = 4.32$, $SD = 1.3$ for Iterative, $M = 4.15$, $SD = 1.63$ for Baseline). This result suggests that in a more complex scenario, humans have a higher likelihood of making mistakes or misunderstanding instructions when delivered all at once, and that by allowing a user to iteratively input preference these mistakes can be reduced. We analogize this to the concept of scaffolding in the learning sciences; when learning more complex concepts, the learner benefits from being presented small pieces one at a time, rather than the whole concept all at once.

5.4.2 User Input

Using the logs that were created when users proceeded between tasks and pages in the experiment, we were able to analyze how much time they spent on the page where they received their new information and provided their input (text and draw annotations on the map). These are detailed in Table 5.3. Because of the nature of this study as a remote online experiment, there were some obvious outliers in this data, suggesting that some users might have ignored the experiment for some time rather than completing all tasks as quickly as possible. Due to these effects, we chose to use median time to more accurately represent the data.

As expected, the Complex Map with slightly lengthier instructions took longer for participants to provide input for than the Simple Map. Furthermore, we also see that the total time spent on either map is somewhat greater in the Iterative condition than in the Baseline condition (30.8% greater for the Complex Map, 15.7% greater for the Simple Map). We saw no significant effect based on which map participants encountered first.

5.4.3 Proving the System

Our system successfully prompted the LLM to choose the most appropriate function for the user's requests and to assign appropriate parameters to that function. Of the 166 total iterations (6 iterations for each participant in the Iterative condition and 2 for each participant in the Baseline condition), the system produced usable output for 157 of them, a 95% success rate. In 6 of the 9 failure occurrences, this was due to users referencing polygons that did not exist (a total of 2 participants). In instances where the LLM failed to provide usable outputs, our system was designed to return the prior trajectory, allowing the experiment to proceed regardless of LLM failures.

We also analyzed how long the trajectory optimization took for each iteration. For the Complex Map, the average duration of the re-planning in the Baseline condition was 86 seconds. For the same map, the average durations in the Iterative condition were 41", 49", and 71" (for each subsequent iteration). For the Simple Map, the Baseline optimization took an average of 65", while the average Iterative durations were 37", 46", and 56". We designed our objective function terms with the goal of completing each optimization within approximately one minute, and these data show how our system successfully prioritized a rapid re-planning to maximize the satisfaction of the human in the loop. This rapid turnaround time reduces the amount of time that a user spent waiting for the system to re-plan, while they completed the somewhat mundane distractor task.

5.4.4 Subjective Measures

While we collected participant responses to Likert questions about their understanding of the system, the predictability of the optimization, and their satisfaction with the final trajectory, we did not find significant differences between the two conditions on these subjective survey measures. However, because those in the Iterative condition ($M = 4.48$, $SD = 1.57$) found the system as equivalently satisfactory as those in the Baseline condition ($M = 4.00$, $SD = 2.00$), this suggests that the requirement to iteratively add preferences to the system, while it did take more time, did not negatively impact their experience. Furthermore, because of the nature of the conditions and execution of the experiment, participants in the Baseline condition completed fewer total distractor tasks (a minimum of 2) than participants in the Iterative condition (a minimum of 6). Despite this, the subjective satisfaction ratings were not significantly different between conditions.

We also obtained some insightful descriptive feedback from participants. Generally, more participants in the Iterative condition claimed that they had an understanding of how to use the system to produce their desired results. Some relevant and representative quotes include: "I understood how the tasks and priority system were effecting [sic] the waypoints," (for the Iterative condition), and "On the last map the trajectory didn't cover second city that should have been searched and I couldn't update the map," (for the Baseline condition). This signals that users preferred iterating on the system output, and that even those not in the Iterative condition asked for this feature.

One participant in the Baseline condition stated that by their second map (which was the Complex Map), they had learned how to write more clear instructions for the system. Likewise, a participant in the Iterative condition pointed out that, "It took me a minute to adjust to how everything worked," but that their second map was "spot on." These examples indicate the clear learning benefits of being able to iteratively interact with the system.

5.5 Summary and Recommendations

To a novice user, and even one with some experience, a system like the one presented here can be perceived as a "black box". Users might have little to no understanding of how the system performs its processing and optimization tasks. However, by allowing users to iteratively add pieces of information and providing visual feedback on that information through display of the subsequent trajectories, the user is afforded some insight into how the system works. This kind of minimal transparency, or "translucency," can aid in training users how best to interact with and use such systems. We show that iteration results in a more preferable optimized trajectory. We also demonstrate that for a more complex scenario, iteration affords users opportunities to learn how to incorporate relevant information (here, in the form of map regions). Ultimately our participants indicated that they liked the ability to iterate on their inputs, and even those who were in the one-shot condition asked for a chance to iterate. We also demonstrated that this system that incorporates the ability to iteratively add latent human knowledge to a trajectory optimization, aided by an LLM-enabled subsystem, and mediated via a web-based user interface, can be used for human-autonomy teaming. In high-risk situations like wildfire search, it is imperative that users be able to iteratively incorporate inputs, improving not only the task outcomes but also user learning and effectiveness.

Map	**Instructions**
Complex	The burned area does not need to be searched as it is low risk.
	The mountains are unnecessary to search because they are above the treeline.
	The areas near and including the towns should be searched. They are high risk because they are highly populated.
Simple	The lake is not at risk for fire and does not need to be searched.
	The burned area does not need to be searched because it is low risk for fire.
	The town in the southwest is also a high fire risk and should be searched.

Table 5.1: The instructions that were presented to participants for each map. Those in the Baseline, one-shot condition received all of the instructions at once for each map. Those in the Iterative condition received one instruction at a time, for a total of 3 iterations.

Function name	Key parameters	Short description
`keep_radius`	polygon, radius	Penalize waypoints for being inside the polygon (or within a given radius)
`explore_area`	polygon	Reward waypoints within the given polygon
`shift_direction`	polygon, direction, distance	Reward waypoints in the polygon for shifting in the given direction

Table 5.2: Summary of information about the additional potential functions provided to the LLM-enabled subsystem. We implemented these functions to be easily added to our objective function for the trajectory optimization.

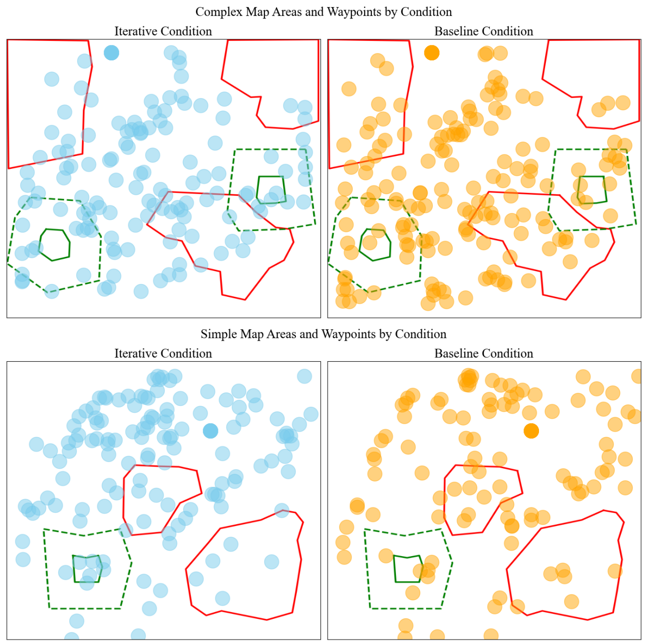


Figure 5.5: Above are the waypoints from all users' final trajectories, separated by condition. Red outlines indicate areas that participants were told to avoid. Green outlines indicate areas that participants were told to search. Positive Waypoint Compliance for the Simple Map was significantly higher for Iterative than for Baseline ($p = .026$). Across both maps, average Waypoint Compliance was significantly higher for Iterative than for the Baseline condition ($p = .010$).

	Baseline	Iterative
Complex	3:04	1:15 (Iteration 1)
		1:35 (Iteration 2)
		1:36 (Iteration 3)
		4:27 (Total)
Simple	2:47	1:19 (Iteration 1)
		1:00 (Iteration 2)
		0:59 (Iteration 3)
		3:18 (Total)

Table 5.3: Median times for users to input their additional information into the interface (text and drawn annotations).

Chapter 6

Conclusions

In the previous chapters I have shown how different emerging technologies can be used to facilitate and mediate human-autonomy teaming. For example, augmented reality can provide a transparent communication modality for humans working in close proximity of robots. And LLM-enabled chatbots are autonomous systems that hold much promise, including for robotics applications.

However, human teammates and users can be predisposed to overtrust these systems or to ignore their instructions. Participants in our study in Chapter 3 failed to comply with the robot's instructions, despite their physical safety being in question. Furthermore, users in our chatbot study disregarded all warnings about the chatbots' fallibility, regardless of being told that they were assessed on accuracy.

In order to seek out ways to improve the human-autonomy team, I designed a system that promoted iteration over one-time interactions. We discovered that not only does iterative disclosure of optimization criteria promote better outcomes, it also results in improved user learning.

6.1 Summary of Findings

Below is a summary of the findings from the research in this dissertation:

- Humans working near robots appear willing to sacrifice some amount of safety to achieve increased efficiency, while also dramatically overestimating the safety of their behavior around the robot.

- During uncertainty, users desire insight into the decisions and priorities of an autonomous system to help them understand the reasoning behind its actions, decrease frustration, and help them make their own decisions about how to act.
- Increasing the perception of a collocated robot's intelligence could significantly decrease a worker's cognitive load.
- In general, even people who say that they do not trust chatbots exhibit overtrust in some situations, making self-proclaimed trust an unreliable signal.
- Even thoughtful, well-informed users are vulnerable to false information, whether inserted intentionally or not.
- Unobtrusive warnings do not appear to alter user behavior.
- Based on prompt length, designers could modify aspects of the chatbot behavior or even refuse to provide an answer.
- By allowing users to iteratively add pieces of information and providing visual feedback on that information through display of the subsequent trajectories, the user is afforded some insight into how the system works.
- "Translucency" can aid in training users how best to interact with and use such systems.
- Iteration achieves better compliance with mission outcomes.
- In complex scenarios, iteration affords users opportunities to learn how to incorporate relevant information.

6.2 Encouraging Responsible Human-Autonomy Collaboration

In addition to publishing research that shares proven practices and recommends appropriate design principles, we can also encourage responsible human-autonomy collaboration by advocating for public policies aligned with research. My work in advance of the proposal of Colorado Senate Bill

22-113 *Artificial Intelligence Facial Recognition*[245], signed into law by the Governor in June 2022, provides one example of this. I also laid outlined guidance not only for involving HRI researchers in public policy, but also for establishing policies within the HRI research community – particularly for increasing transparency around how HRI systems use LLMs [246]. It is important that members of this research field acknowledge their role and their responsibilities when it comes to the power wielded by these technologies.

6.3 Future Work

Each of the works featured in this dissertation provide ample fodder for future research. Questions remain around how we can continue to improve XR-mediated human-robot collaboration, including how to add transparency to the system with visual cues and natural language. LLM-enabled chatbots present a nascent yet rapidly growing collection of questions around ways to require responsible use, both in terms of implementation changes on the back end as well as usage changes on the front end. And human-autonomy teams in the field will continue to require innovations that follow from the findings we presented here around how to onboard users effectively and improve mission outcomes. I can only expect that this field will continue to grow in novel and fascinating ways, and I hope that this work can inform future technological innovations.

Bibliography

[1] P. M. Fitts, Ed., **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, Pages: xxii, 84.

[2] G. Hoffman, "Evaluating Fluency in Human–Robot Collaboration," **IEEE Transactions on Human-Machine Systems**, vol. 49, no. 3, pp. 209–218, Jun. 2019, ISSN: 2168-2291. DOI: 10.1109/THMS.2019.2904558.

[3] C. T. Chang, M. B. Luebbers, M. Hebert, and B. Hayes, "Human non-compliance with robot spatial ownership communicated via augmented reality: Implications for human-robot teaming safety," in **2023 IEEE International Conference on Robotics and Automation (ICRA)**, May 2023, pp. 9785–9792. DOI: 10.1109/ICRA48891.2023.10161277. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10161277 (visited on 10/30/2023).

[4] C. T. Chang, E. Jensen, M. Hebert, and B. Hayes, "Prompt length and search time predict large language model user trust and output verification behaviors," in **In submission**, 2024.

[5] C. T. Chang, M. P. Stull, B. Crockett, E. Jensen, M. Lohrmann Clare Hebert, and B. Hayes, "Iteratively adding latent human knowledge to a trajectory optimization improves learning and task outcomes," in **In submission**, 2024.

[6] P. Milgram, S. Zhai, D. Drascic, and J. Grodski, "Applications of augmented reality for human-robot communication," in **Proceedings of 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '93)**, vol. 3, Jul. 1993, 1467–1472 vol.3. DOI: 10.1109/IROS.1993.583833.

[7] I. Magic Leap, **Magic Leap 1**, en-us, 2018. [Online]. Available: https://www.magicleap.com/magic-leap-1 (visited on 07/14/2020).

[8] microsoft.com, **Microsoft HoloLens — Mixed Reality Technology for Business**, en-us, 2020. [Online]. Available: https://www.microsoft.com/en-us/hololens (visited on 07/14/2020).

[9] S. A. Green, M. Billinghurst, X. Chen, and J. G. Chase, "Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design," en, **International Journal of Advanced Robotic Systems**, vol. 5, no. 1, p. 1, Mar. 2008, Publisher: SAGE Publications, ISSN: 1729-8814. DOI: 10.5772/5664. [Online]. Available: https://doi.org/10.5772/5664 (visited on 05/15/2020).

[10] T. Williams, D. Szafir, T. Chakraborti, and H. Ben Amor, "Virtual, Augmented, and Mixed Reality for Human-Robot Interaction," in **Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI '18, event-place: Chicago, IL, USA, New York, NY, USA: ACM, 2018, pp. 403–404, ISBN: 978-1-4503-5615-2. DOI: 10.1145/3173386.3173561. [Online]. Available: http://doi.acm.org/10.1145/3173386.3173561 (visited on 08/12/2019).

[11] T. Williams, D. Szafir, T. Chakraborti, **et al.**, "Virtual, augmented, and mixed reality for human-robot interaction (vam-hri)," in **Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI '20, Cambridge, United Kingdom: Association for Computing Machinery, 2020, pp. 663–664, ISBN: 9781450370578. DOI: 10.1145/3371382.3374850. [Online]. Available: https://doi.org/10.1145/3371382.3374850.

[12] E. Rosen, T. Groechel, M. E. Walker, C. T. Chang, and J. Z. Forde, "Virtual, augmented, and mixed reality for human-robot interaction (VAM-HRI)," in **Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI '21 Companion, New York, NY, USA: Association for Computing Machinery, Mar. 8, 2021, pp. 721–723, ISBN: 978-1-4503-8290-8. DOI: 10.1145/3434074.3444879. [Online]. Available: https://doi.org/10.1145/3434074.3444879 (visited on 09/19/2021).

[13] C. T. Chang, E. Rosen, T. R. Groechel, M. Walker, and J. Z. Forde, "Virtual, Augmented, and Mixed Reality for HRI (VAM-HRI)," in **Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI '22, Sapporo, Hokkaido, Japan: IEEE Press, Mar. 2022, pp. 1237–1240. (visited on 06/06/2022).

[14] M. Wozniak, C. T. Chang, M. B. Luebbers, **et al.**, "Virtual, augmented, and mixed reality for human-robot interaction (vam-hri)," in **Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI '23, ¡conf-loc¿, ¡city¿Stockholm¡/city¿, ¡country¿Sweden¡/country¿, ¡/conf-loc¿: Association for Computing Machinery, 2023, pp. 938–940, ISBN: 9781450399708. DOI: 10.1145/3568294.3579959. [Online]. Available: https://doi.org/10.1145/3568294.3579959.

[15] T. R. Groechel, M. E. Walker, C. T. Chang, E. Rosen, and J. Z. Forde, "A tool for organizing key characteristics of virtual, augmented, and mixed reality for human–robot interaction systems: Synthesizing VAM-HRI trends and takeaways," **IEEE Robotics & Automation Magazine**, vol. 29, no. 1, pp. 35–44, Mar. 2022, Conference Name: IEEE Robotics & Automation Magazine, ISSN: 1558-223X. DOI: 10.1109/MRA.2021.3138383. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9681717 (visited on 10/30/2023).

[16] C. T. Chang, J. Dixon, M. B. Luebbers, A. Tabrez, and B. Hayes, "Emerging autonomy solutions for human and robotic deep space exploration," in **Proceedings of Workshop**

on SpaceCHI : Human-Computer Interaction for Space Exploration (SpaceCHI '21), Virtual: ACM, 2021, p. 5.

[17] J. Chestnutt, K. Nishiwaki, J. Kuffner, and S. Kagamiy, "Interactive control of humanoid navigation," in **2009 IEEE/RSJ International Conference on Intelligent Robots and Systems**, ISSN: 2153-0866, Oct. 2009, pp. 3519–3524. DOI: 10.1109/IROS.2009.5354571.

[18] M. Walker, H. Hedayati, J. Lee, and D. Szafir, "Communicating robot motion intent with augmented reality," in **Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI '18, event-place: Chicago, IL, USA, New York, NY, USA: ACM, 2018, pp. 316–324, ISBN: 978-1-4503-4953-6. DOI: 10.1145/3171221.3171253. [Online]. Available: http://doi.acm.org/10.1145/3171221.3171253 (visited on 06/26/2019).

[19] M. Zolotas and Y. Demiris, "Towards Explainable Shared Control using Augmented Reality," in **2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Nov. 2019, pp. 3020–3026. DOI: 10.1109/IROS40897.2019.8968117.

[20] M. E. Walker, H. Hedayati, and D. Szafir, "Robot teleoperation with augmented reality virtual surrogates," in **Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI '19, Daegu, Republic of Korea: IEEE Press, Mar. 2019, pp. 202–210, ISBN: 978-1-5386-8555-6. (visited on 05/21/2020).

[21] S. A. Green, J. G. Chase, X. Chen, and M. Billinghurst, "Evaluating the augmented reality human-robot collaboration system," **International Journal of Intelligent Systems Technologies and Applications**, vol. 8, no. 1-4, pp. 130–143, Dec. 2009, Publisher: Inderscience Publishers, ISSN: 1740-8865. DOI: 10.1504/IJISTA.2010.030195. [Online]. Available: https://www.inderscienceonline.com/doi/abs/10.1504/IJISTA.2010.030195 (visited on 05/15/2020).

[22] E. Oyama, N. Shiroma, M. Niwa, **et al.**, "Hybrid head mounted/surround display for telexistence/telepresence and behavior navigation," in **2013 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)**, ISSN: 2374-3247, Oct. 2013, pp. 1–6. DOI: 10.1109/SSRR.2013.6719352.

[23] K. Krückel, F. Nolden, A. Ferrein, and I. Scholl, "Intuitive visual teleoperation for UGVs using free-look augmented reality displays," in **2015 IEEE International Conference on Robotics and Automation (ICRA)**, ISSN: 1050-4729, May 2015, pp. 4412–4417. DOI: 10.1109/ICRA.2015.7139809.

[24] J. Guhl, S. Tung, and J. Kruger, "Concept and architecture for programming industrial robots using augmented reality with mobile devices like microsoft HoloLens," in **2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)**, ISSN: 1946-0759, Sep. 2017, pp. 1–4. DOI: 10.1109/ETFA.2017.8247749.

[25] A. W. W. Yew, S. K. Ong, and A. Y. C. Nee, "Immersive Augmented Reality Environment for the Teleoperation of Maintenance Robots," en, **Procedia CIRP**, The 24th CIRP Conference on Life Cycle Engineering, vol. 61, pp. 305–310, Jan. 2017, ISSN: 2212-8271. DOI: 10.1016/j.procir.2016.11.183. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2212827116313439 (visited on 05/19/2020).

[26] M. Zolotas, J. Elsdon, and Y. Demiris, "Head-Mounted Augmented Reality for Explainable Robotic Wheelchair Assistance," in **2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Oct. 2018, pp. 1823–1829. DOI: 10.1109/IROS.2018.8594002.

[27] R. Chacón-Quesada and Y. Demiris, "Augmented Reality Controlled Smart Wheelchair Using Dynamic Signifiers for Affordance Representation," in **2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Nov. 2019, pp. 4812–4818. DOI: 10.1109/IROS40897.2019.8968290.

[28] M. Rudorfer, J. Guhl, P. Hoffmann, and J. Krüger, "Holo Pick'n'Place," in **2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)**, ISSN: 1946-0759, vol. 1, Sep. 2018, pp. 1219–1222. DOI: 10.1109/ETFA.2018.8502527.

[29] D. Puljiz, E. Stöhr, K. S. Riesterer, B. Hein, and T. Kröger, "General Hand Guidance Framework using Microsoft HoloLens," in **2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Nov. 2019, pp. 5185–5190. DOI: 10.1109/IROS40897.2019.8967649.

[30] J. Elsdon and Y. Demiris, "Augmented Reality for Feedback in a Shared Control Spraying Task," in **2018 IEEE International Conference on Robotics and Automation (ICRA)**, ISSN: 2577-087X, May 2018, pp. 1939–1946. DOI: 10.1109/ICRA.2018.8461179.

[31] C. Reardon, K. Lee, and J. Fink, "Come See This! Augmented Reality to Enable Human-Robot Cooperative Search," in **2018 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)**, ISSN: 2475-8426, Aug. 2018, pp. 1–7. DOI: 10.1109/SSRR.2018.8468622.

[32] L. Kästner and J. Lambrecht, "Augmented-Reality-Based Visualization of Navigation Data of Mobile Robots on the Microsoft Hololens - Possibilities and Limitations," in **2019 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM)**, ISSN: 2326-8239, Nov. 2019, pp. 344–349. DOI: 10.1109/CIS-RAM47153.2019.9095836.

[33] H. Hedayati, M. Walker, and D. Szafir, "Improving Collocated Robot Teleoperation with Augmented Reality," in **Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI '18, Chicago, IL, USA: Association for Computing Machinery, Feb. 2018, pp. 78–86, ISBN: 978-1-4503-4953-6. DOI: 10.1145/3171221.3171251. [Online]. Available: https://doi.org/10.1145/3171221.3171251 (visited on 05/15/2020).

[34] L. Qian, A. Deguet, Z. Wang, Y.-H. Liu, and P. Kazanzides, "Augmented Reality Assisted Instrument Insertion and Tool Manipulation for the First Assistant in Robotic Surgery," in **2019 International Conference on Robotics and Automation (ICRA)**, ISSN: 2577-087X, May 2019, pp. 5173–5179. DOI: 10.1109/ICRA.2019.8794263.

[35] R. Fung, S. Hashimoto, M. Inami, and T. Igarashi, "An augmented reality system for teaching sequential tasks to a household robot," in **2011 RO-MAN**, ISSN: 1944-9437, Jul. 2011, pp. 282–287. DOI: 10.1109/ROMAN.2011.6005235.

[36] J. Lambrecht and J. Krüger, "Spatial programming for industrial robots based on gestures and Augmented Reality," in **2012 IEEE/RSJ International Conference on Intelligent Robots and Systems**, ISSN: 2153-0866, Oct. 2012, pp. 466–472. DOI: 10.1109/IROS.2012.6385900.

[37] S. Bonardi, J. Blatter, J. Fink, **et al.**, "Design and evaluation of a graphical iPad application for arranging adaptive furniture," in **2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication**, ISSN: 1944-9437, Sep. 2012, pp. 290–297. DOI: 10.1109/ROMAN.2012.6343768.

[38] S. Stadler, K. Kain, M. Giuliani, N. Mirnig, G. Stollnberger, and M. Tscheligi, "Augmented reality for industrial robot programmers: Workload analysis for task-based, augmented reality-supported robot control," in **2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)**, ISSN: 1944-9437, Aug. 2016, pp. 179–184. DOI: 10.1109/ROMAN.2016.7745108.

[39] J. Hügle, J. Lambrecht, and J. Krüger, "An integrated approach for industrial robot control and programming combining haptic and non-haptic gestures," in **2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)**, ISSN: 1944-9437, Aug. 2017, pp. 851–857. DOI: 10.1109/ROMAN.2017.8172402.

[40] J. A. Frank, M. Moorhead, and V. Kapila, "Mobile Mixed-Reality Interfaces That Enhance Human–Robot Interaction in Shared Spaces," English, **Frontiers in Robotics and AI**,

vol. 4, 2017, Publisher: Frontiers, ISSN: 2296-9144. DOI: 10.3389/frobt.2017.00020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frobt.2017.00020/full (visited on 05/15/2020).

[41] D. Sprute, K. Tönnies, and M. König, "Virtual Borders: Accurate Definition of a Mobile Robot's Workspace Using Augmented Reality," in **2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Oct. 2018, pp. 8574–8581. DOI: 10.1109/IROS.2018.8593615.

[42] S. M. Chacko and V. Kapila, "Augmented Reality as a Medium for Human-Robot Collaborative Tasks," in **2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)**, ISSN: 1944-9437, Oct. 2019, pp. 1–8. DOI: 10.1109/RO-MAN46459.2019.8956466.

[43] A. Rotsidis, A. Theodorou, J. J. Bryson, and R. H. Wortham, "Improving Robot Transparency: An Investigation With Mobile Augmented Reality," in **2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)**, ISSN: 1944-9437, Oct. 2019, pp. 1–8. DOI: 10.1109/RO-MAN46459.2019.8956390.

[44] R. S. Andersen, S. Bøgh, T. B. Moeslund, and O. Madsen, "Task space HRI for cooperative mobile robots in fit-out operations inside ship superstructures," in **2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)**, ISSN: 1944-9437, Aug. 2016, pp. 880–887. DOI: 10.1109/ROMAN.2016.7745223.

[45] R. Kalpagam Ganesan, Y. K. Rathore, H. M. Ross, and H. Ben Amor, "Better Teaming Through Visual Cues: How Projecting Imagery in a Workspace Can Improve Human-Robot Collaboration," en, **IEEE Robotics & Automation Magazine**, vol. 25, no. 2, pp. 59–71, Jun. 2018, ISSN: 1070-9932, 1558-223X. DOI: 10.1109/MRA.2018.2815655. [Online]. Available: https://ieeexplore.ieee.org/document/8359206/ (visited on 05/15/2020).

[46] Z. Materna, M. Kapinus, V. Beran, P. Smrž, and P. Zemčík, "Interactive Spatial Augmented Reality in Collaborative Robot Programming: User Experience Evaluation," in **2018 27th**

IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), ISSN: 1944-9437, Aug. 2018, pp. 80–87. DOI: 10.1109/ROMAN.2018.8525662.

[47] G. Bolano, C. Juelg, A. Roennau, and R. Dillmann, "Transparent Robot Behavior Using Augmented Reality in Close Human-Robot Interaction," in **2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)**, ISSN: 1944-9437, Oct. 2019, pp. 1–7. DOI: 10.1109/RO-MAN46459.2019.8956296.

[48] T. Ito, T. Niwa, and A. H. Slocum, "Virtual cutter path display for dental milling machine," in **RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication**, ISSN: 1944-9437, Sep. 2009, pp. 488–493. DOI: 10.1109/ROMAN.2009.5326247.

[49] S. Notheis, G. Milighetti, B. Hein, H. Wörn, and J. Beyerer, "Skill-based telemanipulation by means of intelligent robots," in **2010 IEEE/RSJ International Conference on Intelligent Robots and Systems**, ISSN: 2153-0866, Oct. 2010, pp. 5258–5263. DOI: 10.1109/IROS.2010.5650731.

[50] C. Domingues, M. Essabbah, N. Cheaib, S. Otmane, and A. Dinis, "Human-Robot-Interfaces based on Mixed Reality for Underwater Robot Teleoperation," en, **IFAC Proceedings Volumes**, 9th IFAC Conference on Manoeuvring and Control of Marine Craft, vol. 45, no. 27, pp. 212–215, Jan. 2012, ISSN: 1474-6670. DOI: 10.3182/20120919-3-IT-2046.00036. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474667016312307 (visited on 05/29/2020).

[51] S. Hashimoto, A. Ishida, M. Inami, and T. Igarashi, "TouchMe: An Augmented Reality Interface for Remote Robot Control," **Journal of Robotics and Mechatronics**, vol. 25, no. 3, pp. 529–537, Jun. 2013, Publisher: Fuji Technology Press Ltd. DOI: 10.20965/jrm.2013.p0529. [Online]. Available: https://www.fujipress.jp/jrm/rb/robot002500030529/ (visited on 05/29/2020).

[52] A. Osaki, T. Kaneko, and Y. Miwa, “Embodied navigation for mobile robot by using direct 3D drawing in the air,” in **RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication**, ISSN: 1944-9437, Aug. 2008, pp. 671–676. DOI: 10.1109/ROMAN.2008.4600744.

[53] F.-J. Chu, R. Xu, Z. Zhang, P. A. Vela, and M. Ghovanloo, “Hands-Free Assistive Manipulator Using Augmented Reality and Tongue Drive System,” in **2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Oct. 2018, pp. 5463–5468. DOI: 10.1109/IROS.2018.8594508.

[54] S. Oota, A. Murai, and M. Mochimaru, “Lucid Virtual/Augmented Reality (LVAR) Integrated with an Endoskeletal Robot Suit: StillSuit: A new framework for cognitive and physical interventions to support the ageing society,” in **2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)**, ISSN: 2642-5254, Mar. 2019, pp. 1556–1559. DOI: 10.1109/VR.2019.8798012.

[55] J. M. Gregory, C. Reardon, K. Lee, G. White, K. Ng, and C. Sims, “Enabling Intuitive Human-Robot Teaming Using Augmented Reality and Gesture Control,” **arXiv:1909.06415 [cs]**, Sep. 2019, arXiv: 1909.06415. [Online]. Available: http://arxiv.org/abs/1909.06415 (visited on 09/25/2020).

[56] D. Q. Huy, I. Vietcheslav, and G. Seet Gim Lee, “See-through and spatial augmented reality - a novel framework for human-robot interaction,” in **2017 3rd International Conference on Control, Automation and Robotics (ICCAR)**, Apr. 2017, pp. 719–726. DOI: 10.1109/ICCAR.2017.7942791.

[57] E. Sibirtseva, D. Kontogiorgos, O. Nykvist, **et al.**, “A Comparison of Visualisation Methods for Disambiguating Verbal Requests in Human-Robot Interaction,” in **2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)**, ISSN: 1944-9437, Aug. 2018, pp. 43–50. DOI: 10.1109/ROMAN.2018.8525554.

[58] D. Bambušek, Z. Materna, M. Kapinus, V. Beran, and P. Smrž, “Combining Interactive Spatial Augmented Reality with Head-Mounted Display for End-User Collaborative Robot Programming,” in **2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)**, ISSN: 1944-9437, Oct. 2019, pp. 1–8. DOI: 10.1109/RO-MAN46459.2019.8956315.

[59] D. Sportillo, A. Paljic, and L. Ojeda, “On-road evaluation of autonomous driving training,” in **Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI ’19, Daegu, Republic of Korea: IEEE Press, Mar. 2019, pp. 182–190, ISBN: 978-1-5386-8555-6. (visited on 05/21/2020).

[60] T. Chakraborti, S. Sreedharan, A. Kulkarni, and S. Kambhampati, “Projection-Aware Task Planning and Execution for Human-in-the-Loop Operation of Robots in a Mixed-Reality Workspace,” in **2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Oct. 2018, pp. 4476–4482. DOI: 10.1109/IROS.2018.8593830.

[61] D. Sprute, P. Viertel, K. Tönnies, and M. König, “Learning Virtual Borders through Semantic Scene Understanding and Augmented Reality,” in **2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Nov. 2019, pp. 4607–4614. DOI: 10.1109/IROS40897.2019.8967576.

[62] C. Reardon, K. Lee, J. G. Rogers, and J. Fink, “Communicating via Augmented Reality for Human-Robot Teaming in Field Environments,” in **2019 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)**, ISSN: 2475-8426, Sep. 2019, pp. 94–101. DOI: 10.1109/SSRR.2019.8848971.

[63] T. Williams, M. Bussing, S. Cabrol, E. Boyle, and N. Tran, “Mixed reality deictic gesture for multi-modal robot communication,” in **2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)**, ISSN: 2167-2148, Mar. 2019, pp. 191–201. DOI: 10.1109/HRI.2019.8673275.

[64] J. Hamilton, T. Phung, N. Tran, and T. Williams, “What’s The Point? Tradeoffs Between Effectiveness and Social Perception When Using Mixed Reality to Enhance Gesturally Limited Robots,” in **Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI ’21, New York, NY, USA: Association for Computing Machinery, Mar. 2021, pp. 177–186, ISBN: 978-1-4503-8289-2. DOI: 10.1145/3434073.3444676. [Online]. Available: https://doi.org/10.1145/3434073.3444676 (visited on 10/20/2021).

[65] K. Chandan, V. Kudalkar, X. Li, and S. Zhang, “ARROCH: Augmented Reality for Robots Collaborating with a Human,” in **2021 IEEE International Conference on Robotics and Automation (ICRA)**, ISSN: 2577-087X, May 2021, pp. 3787–3793. DOI: 10.1109/ICRA48506.2021.9561144.

[66] B. Ikeda and D. Szafir, “Advancing the Design of Visual Debugging Tools for Roboticists,” in **Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI ’22, Sapporo, Hokkaido, Japan: IEEE Press, Mar. 2022, pp. 195–204. (visited on 08/10/2022).

[67] G. Reinhart, U. Munzert, and W. Vogl, “A programming system for robot-based remote-laser-welding with conventional optics,” en, **CIRP Annals**, vol. 57, no. 1, pp. 37–40, Jan. 2008, ISSN: 0007-8506. DOI: 10.1016/j.cirp.2008.03.120. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0007850608000917 (visited on 05/29/2020).

[68] T. Hulin, V. Schmirgel, E. Yechiam, U. E. Zimmermann, C. Preusche, and G. Pöhler, “Evaluating exemplary training accelerators for Programming-by-Demonstration,” in **19th International Symposium in Robot and Human Interactive Communication**, ISSN: 1944-9437, Sep. 2010, pp. 440–445. DOI: 10.1109/ROMAN.2010.5598611.

[69] M. Gianni, G. Gonnelli, A. Sinha, M. Menna, and F. Pirri, “An Augmented Reality approach for trajectory planning and control of tracked vehicles in rescue environments,” in **2013**

IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), ISSN: 2374-3247, Oct. 2013, pp. 1–6. DOI: 10.1109/SSRR.2013.6719360.

[70] J. Lambrecht, H. Walzel, and J. Krüger, "Robust finger gesture recognition on handheld devices for spatial programming of industrial robots," in **2013 IEEE RO-MAN**, ISSN: 1944-9437, Aug. 2013, pp. 99–106. DOI: 10.1109/ROMAN.2013.6628462.

[71] M. D. Coovert, T. Lee, I. Shindev, and Y. Sun, "Spatial augmented reality as a method for a mobile robot to communicate intended movement," en, **Computers in Human Behavior**, vol. 34, pp. 241–248, May 2014, ISSN: 0747-5632. DOI: 10.1016/j.chb.2014.02.001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563214000612 (visited on 09/25/2020).

[72] R. T. Chadalavada, H. Andreasson, R. Krug, and A. J. Lilienthal, "That's on my mind! robot to human intention communication through on-board projection on shared floor space," in **2015 European Conference on Mobile Robots (ECMR)**, Sep. 2015, pp. 1–6. DOI: 10.1109/ECMR.2015.7403771.

[73] S. Makris, P. Karagiannis, S. Koukas, and A.-S. Matthaiakis, "Augmented reality system for operator support in human–robot collaborative assembly," **CIRP Annals**, vol. 65, no. 1, pp. 61–64, Jan. 1, 2016, ISSN: 0007-8506. DOI: 10.1016/j.cirp.2016.04.038. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0007850616300385 (visited on 05/29/2020).

[74] D. Krupke, F. Steinicke, P. Lubos, Y. Jonetzko, M. Görner, and J. Zhang, "Comparison of Multimodal Heading and Pointing Gestures for Co-Located Mixed Reality Human-Robot Interaction," in **2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Oct. 2018, pp. 1–9. DOI: 10.1109/IROS.2018.8594043.

[75] M. Kapinus, V. Beran, Z. Materna, and D. Bambušek, "Spatially Situated End-User Robot Programming in Augmented Reality," in **2019 28th IEEE International Conference on**

Robot and Human Interactive Communication (RO-MAN), ISSN: 1944-9437, Oct. 2019, pp. 1–8. DOI: 10.1109/RO-MAN46459.2019.8956336.

[76] C. Liu and S. Shen, “An Augmented Reality Interaction Interface for Autonomous Drone,” in **arXiv:2008.02234 [cs]**, arXiv: 2008.02234, Aug. 2020. [Online]. Available: http://arxiv.org/abs/2008.02234 (visited on 09/09/2020).

[77] A. Corotan and J. J. Z. Irgen-Gioro, “An Indoor Navigation Robot Using Augmented Reality,” in **2019 5th International Conference on Control, Automation and Robotics (ICCAR)**, ISSN: 2251-2446, Apr. 2019, pp. 111–116. DOI: 10.1109/ICCAR.2019.8813348.

[78] S. Y. Gadre, E. Rosen, G. Chien, E. Phillips, S. Tellex, and G. Konidaris, “End-User Robot Programming Using Mixed Reality,” in **2019 International Conference on Robotics and Automation (ICRA)**, ISSN: 2577-087X, May 2019, pp. 2707–2713. DOI: 10.1109/ICRA.2019.8793988.

[79] M. Ostanin, S. Mikhel, A. Evlampiev, V. Skvortsova, and A. Klimchik, “Human-robot interaction for robotic manipulator programming in Mixed Reality,” in **2020 IEEE International Conference on Robotics and Automation (ICRA)**, ISSN: 2577-087X, May 2020, pp. 2805–2811. DOI: 10.1109/ICRA40945.2020.9196965.

[80] M. B. Luebbers, C. Brooks, C. L. Mueller, D. Szafir, and B. Hayes, “ARC-LfD: Using Augmented Reality for Interactive Long-Term Robot Skill Maintenance via Constrained Learning from Demonstration,” in **2021 IEEE International Conference on Robotics and Automation (ICRA)**, ISSN: 2577-087X, May 2021, pp. 3794–3800. DOI: 10.1109/ICRA48506.2021.9561844.

[81] Z. Han, J. Parrillo, A. Wilkinson, H. A. Yanco, and T. Williams, “Projecting Robot Navigation Paths: Hardware and Software for Projected AR,” in **Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI ’22, Sapporo, Hokkaido, Japan: IEEE Press, Mar. 2022, pp. 623–628. (visited on 08/10/2022).

[82] S. A. Green, X. Q. Chen, M. Billinghurst, and J. G. Chase, "Collaborating with a Mobile Robot: An Augmented Reality Multimodal Interface," en, **IFAC Proceedings Volumes**, 17th IFAC World Congress, vol. 41, no. 2, pp. 15 595–15 600, Jan. 2008, ISSN: 1474-6670. DOI: 10.3182/20080706-5-KR-1001.02637. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474667016415028 (visited on 05/29/2020).

[83] W. Hönig, C. Milanes, L. Scaria, T. Phan, M. Bolas, and N. Ayanian, "Mixed reality for robotics," in **2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, Sep. 2015, pp. 5382–5387. DOI: 10.1109/IROS.2015.7354138.

[84] I. D. Peake, J. O. Blech, and M. Schembri, "A software framework for augmented reality-based support of industrial operations," in **2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)**, Sep. 2016, pp. 1–4. DOI: 10.1109/ETFA.2016.7733627.

[85] R. S. Andersen, O. Madsen, T. B. Moeslund, and H. B. Amor, "Projecting robot intentions into human environments," in **2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)**, ISSN: 1944-9437, Aug. 2016, pp. 294–301. DOI: 10.1109/ROMAN.2016.7745145.

[86] D. Puljiz, F. Krebs, F. Bösing, and B. Hein, "What the HoloLens Maps Is Your Workspace: Fast Mapping and Set-up of Robot Cells via Head Mounted Displays and Augmented Reality," in **arXiv:2005.12651 [cs]**, arXiv: 2005.12651, May 2020. [Online]. Available: http://arxiv.org/abs/2005.12651 (visited on 09/09/2020).

[87] J. T. Hing, K. W. Sevcik, and P. Y. Oh, "Improving unmanned aerial vehicle pilot training and operation for flying in cluttered environments," in **2009 IEEE/RSJ International Conference on Intelligent Robots and Systems**, ISSN: 2153-0866, Oct. 2009, pp. 5641–5646. DOI: 10.1109/IROS.2009.5354080.

[88] J. Riordan, J. Horgan, and D. Toal, "A Real-Time Subsea Environment Visualisation Framework for Simulation of Vision Based UUV Control Architectures," en, **IFAC Proceedings**

Volumes, 2nd IFAC Workshop on Navigation, Guidance and Control of Underwater Vehicles, vol. 41, no. 1, pp. 25–30, Jan. 2008, ISSN: 1474-6670. DOI: 10.3182/20080408-3-IE-4914.00006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474667015354963 (visited on 05/29/2020).

[89] C. Brooks and D. Szafir, "Visualization of Intended Assistance for Acceptance of Shared Control," in **2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Oct. 2020, pp. 11 425–11 430. DOI: 10.1109/IROS45743.2020.9340964.

[90] S. O. Sachidanandam, S. Honarvar, and Y. Diaz-Mercado, "Effectiveness of Augmented Reality for Human Swarm Interactions," in **2022 International Conference on Robotics and Automation (ICRA)**, May 2022, pp. 11 258–11 264. DOI: 10.1109/ICRA46639.2022.9812080.

[91] S. M. Chacko and V. Kapila, "Augmented Reality as a Medium for Human-Robot Collaborative Tasks," in **2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)**, ISSN: 1944-9437, Oct. 2019, pp. 1–8. DOI: 10.1109/RO-MAN46459.2019.8956466.

[92] H. Martins and R. Ventura, "Immersive 3-D teleoperation of a search and rescue robot using a head-mounted display," in **2009 IEEE Conference on Emerging Technologies Factory Automation**, ISSN: 1946-0759, Sep. 2009, pp. 1–8. DOI: 10.1109/ETFA.2009.5347014.

[93] L. Zalud, P. Kocmanova, F. Burian, and T. Jilek, "Color and Thermal Image Fusion for Augmented Reality in Rescue Robotics," en, in **The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications**, H. A. Mat Sakim and M. T. Mustaffa, Eds., ser. Lecture Notes in Electrical Engineering, Singapore: Springer, 2014, pp. 47–55, ISBN: 978-981-4585-42-2. DOI: 10.1007/978-981-4585-42-2_6.

[94] C. Reardon, K. Haring, J. M. Gregory, and J. G. Rogers, "Evaluating Human Understanding of a Mixed Reality Interface for Autonomous Robot-Based Change Detection," in **2021 IEEE**

International Symposium on Safety, Security, and Rescue Robotics (SSRR), ISSN: 2475-8426, Oct. 2021, pp. 132–137. DOI: 10.1109/SSRR53300.2021.9597854.

[95] M. Walker, Z. Chen, M. Whitlock, **et al.**, "A Mixed Reality Supervision and Telepresence Interface for Outdoor Field Robotics," in **2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Sep. 2021, pp. 2345–2352. DOI: 10.1109/IROS51168.2021.9636090.

[96] A. Tabrez, M. B. Luebbers, and B. Hayes, "Descriptive and prescriptive visual guidance to improve shared situational awareness in human-robot teaming," in **Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems**, 2022, pp. 1256–1264.

[97] L. Qian, J. Y. Wu, S. P. DiMaio, N. Navab, and P. Kazanzides, "A Review of Augmented Reality in Robotic-Assisted Surgery," **IEEE Transactions on Medical Robotics and Bionics**, vol. 2, no. 1, pp. 1–16, Feb. 2020, Conference Name: IEEE Transactions on Medical Robotics and Bionics, ISSN: 2576-3202. DOI: 10.1109/TMRB.2019.2957061.

[98] E. Oyama, N. Watanabe, H. Mikado, **et al.**, "A study on wearable behavior navigation system (II) - a comparative study on remote behavior navigation systems for first-aid treatment," in **19th International Symposium in Robot and Human Interactive Communication**, ISSN: 1944-9437, Sep. 2010, pp. 755–761. DOI: 10.1109/ROMAN.2010.5598655.

[99] A. Filippeschi, F. Brizzi, E. Ruffaldi, J. M. Jacinto, and C. A. Avizzano, "Encountered-type haptic interface for virtual interaction with real objects based on implicit surface haptic rendering for remote palpation," in **2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, Sep. 2015, pp. 5904–5909. DOI: 10.1109/IROS.2015.7354216.

[100] Y. Adagolodjo, R. Trivisonne, N. Haouchine, S. Cotin, and H. Courtecuisse, "Silhouette-based pose estimation for deformable organs application to surgical augmented reality," in **2017**

IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), ISSN: 2153-0866, Sep. 2017, pp. 539–544. DOI: `10.1109/IROS.2017.8202205`.

[101] N. Zevallos, A. S. Rangaprasad, H. Salman, **et al.**, "A Real-time Augmented Reality Surgical System for Overlaying Stiffness Information," vol. 14, Jun. 2018, ISBN: 978-0-9923747-4-7. [Online]. Available: `http://www.roboticsproceedings.org/rss14/p26.html` (visited on 05/26/2020).

[102] T. Sheridan, "Space teleoperation through time delay: Review and prognosis," **IEEE Transactions on Robotics and Automation**, vol. 9, pp. 592–606, Oct. 1993, Conference Name: IEEE Transactions on Robotics and Automation, ISSN: 2374-958X. DOI: `10.1109/70.258052`.

[103] T. Xia, S. Léonard, A. Deguet, L. Whitcomb, and P. Kazanzides, "Augmented reality environment with virtual fixtures for robotic telemanipulation in space," in **2012 IEEE/RSJ International Conference on Intelligent Robots and Systems**, ISSN: 2153-0866, Oct. 2012, pp. 5059–5064. DOI: `10.1109/IROS.2012.6386169`.

[104] H. Ro, J.-H. Byun, I. Kim, Y. J. Park, K. Kim, and T.-D. Han, "Projection-Based Augmented Reality Robot Prototype with Human-Awareness," in **2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)**, ISSN: 2167-2148, Mar. 2019, pp. 598–599. DOI: `10.1109/HRI.2019.8673173`.

[105] N. Mavridis and D. Hanson, "The IbnSina Center: An augmented reality theater with intelligent robotic and virtual characters," in **RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication**, ISSN: 1944-9437, Sep. 2009, pp. 681–686. DOI: `10.1109/ROMAN.2009.5326148`.

[106] A. Pereira, E. J. Carter, I. Leite, J. Mars, and J. F. Lehman, "Augmented reality dialog interface for multimodal teleoperation," in **2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)**, ISSN: 1944-9437, Aug. 2017, pp. 764–771. DOI: `10.1109/ROMAN.2017.8172389`.

[107] S. Omidshafiei, A. Agha-Mohammadi, Y. F. Chen, **et al.**, "Measurable Augmented Reality for Prototyping Cyberphysical Systems: A Robotics Platform to Aid the Hardware Prototyping and Performance Testing of Algorithms," **IEEE Control Systems Magazine**, vol. 36, no. 6, pp. 65–87, Dec. 2016, Conference Name: IEEE Control Systems Magazine, ISSN: 1941-000X. DOI: 10.1109/MCS.2016.2602090.

[108] K. Mahajan, T. R. Groechel, R. Pakkar, H. J. Lee, J. Cordero, and M. J. Matarić, "Adapting Usability Metrics for a Socially Assistive, Kinesthetic, Mixed Reality Robot Tutoring Environment," in **International Conference on Social Robotics**, Nov. 2020. [Online]. Available: https://robotics.usc.edu/publications/1107/ (visited on 10/07/2020).

[109] S. M. Chacko and V. Kapila, "An Augmented Reality Interface for Human-Robot Interaction in Unconstrained Environments," in **2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Nov. 2019, pp. 3222–3228. DOI: 10.1109/IROS40897.2019.8967973.

[110] J. C. R. Licklider, "Man-Computer Symbiosis," **IRE Transactions on Human Factors in Electronics**, vol. HFE-1, no. 1, pp. 4–11, Mar. 1960, Conference Name: IRE Transactions on Human Factors in Electronics, ISSN: 2168-2836. DOI: 10.1109/THFE2.1960.4503259.

[111] T. B. Sheridan, **Telerobotics, Automation, and Human Supervisory Control**, en. Cambridge, MA, USA: MIT Press, Aug. 1992, ISBN: 978-0-262-19316-0.

[112] C. Wickens, G. Dempsey, A. Pringle, L. Kazansky, and S. Hutka, "Developing and evaluating an augmented reality interface to assist the joint tactical air controller by applying human performance models," in **Proceedings of the Human Factors and Ergonomics Society Annual Meeting**, SAGE Publications Sage CA: Los Angeles, CA, vol. 62, 2018, pp. 686–690.

[113] NASA Ames, **Nasa ames**, Aug. 2019. [Online]. Available: https://humansystems.arc.nasa.gov/groups/TLX/ (visited on 08/25/2020).

[114] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots,"

en, **International Journal of Social Robotics**, vol. 1, no. 1, pp. 71–81, Jan. 2009, ISSN: 1875-4791, 1875-4805. DOI: 10.1007/s12369-008-0001-3. [Online]. Available: http://link.springer.com/10.1007/s12369-008-0001-3 (visited on 08/25/2020).

[115] A. Weiss and C. Bartneck, "Meta analysis of the usage of the Godspeed Questionnaire Series," in **2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)**, Aug. 2015, pp. 381–388. DOI: 10.1109/ROMAN.2015.7333568.

[116] B. Laugwitz, T. Held, and M. Schrepp, "Construction and Evaluation of a User Experience Questionnaire," en, in **HCI and Usability for Education and Work**, A. Holzinger, Ed., ser. Lecture Notes in Computer Science, Berlin, Heidelberg: Springer, 2008, pp. 63–76, ISBN: 978-3-540-89350-9. DOI: 10.1007/978-3-540-89350-9_6.

[117] J. Brooke, "Usability and Context," en, in **Usability Evaluation In Industry**, P. W. Jordan, B. Thomas, I. L. McClelland, and B. Weerdmeester, Eds., Google-Books-ID: ujFRDwAAQBAJ, CRC Press, Jun. 1996, ISBN: 978-1-4987-1041-1.

[118] M. Endsley, "Situation awareness global assessment technique (SAGAT)," in **Proceedings of the IEEE 1988 National Aerospace and Electronics Conference**, May 1988, 789–795 vol.3. DOI: 10.1109/NAECON.1988.195097.

[119] D. U. Wheelchair Skills Program, **Wheelchair Skills Program (WSP) Version 4.2 – Wheelchair Skills Program**, en-CA, 2013. [Online]. Available: https://wheelchairskillsprogram.ca/en/skills-manual-forms-version-4-2/ (visited on 09/01/2020).

[120] R. Kirby, J. Swuste, D. J. Dupuis, D. A. MacLeod, and R. Monroe, "The Wheelchair Skills Test: A pilot study of a new outcome measure," en, **Archives of Physical Medicine and Rehabilitation**, vol. 83, no. 1, pp. 10–18, Jan. 2002, ISSN: 00039993. DOI: 10.1053/apmr.2002.26823. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0003999302899956 (visited on 09/01/2020).

[121] G. Hoffman, **Evaluating fluency in human-robot collaboration**, Berlin, Germany, 2013.

[122] M. C. Gombolay, R. A. Gutierrez, S. G. Clarke, G. F. Sturla, and J. A. Shah, "Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams," en, **Autonomous Robots**, vol. 39, no. 3, pp. 293–312, Oct. 2015, ISSN: 0929-5593, 1573-7527. DOI: 10.1007/s10514-015-9457-9. [Online]. Available: http://link.springer.com/10.1007/s10514-015-9457-9 (visited on 03/02/2020).

[123] A. D. Dragan, S. Bauman, J. Forlizzi, and S. S. Srinivasa, "Effects of Robot Motion on Human-Robot Collaboration," in **Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI '15, event-place: Portland, Oregon, USA, New York, NY, USA: ACM, 2015, pp. 51–58, ISBN: 978-1-4503-2883-8. DOI: 10.1145/2696454.2696473. [Online]. Available: http://doi.acm.org/10.1145/2696454.2696473 (visited on 05/28/2019).

[124] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," en, in **Advances in Psychology**, ser. Human Mental Workload, P. A. Hancock and N. Meshkati, Eds., vol. 52, North-Holland, Jan. 1988, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9. (visited on 08/26/2020).

[125] L. Srinivasan and K. Schilling, "Augmented Reality Exocentric Navigation Paradigm for Time Delayed Teleoperation," en, **IFAC Proceedings Volumes**, 3rd IFAC Symposium on Telematics Applications, vol. 46, no. 29, pp. 1–6, Jan. 2013, ISSN: 1474-6670. DOI: 10.3182/20131111-3-KR-2043.00002. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1474667015343524 (visited on 05/29/2020).

[126] D. Szafir, B. Mutlu, and T. Fong, "Designing planning and control interfaces to support user collaboration with flying robots," en, **The International Journal of Robotics Research**, vol. 36, no. 5-7, pp. 514–542, Jun. 2017, Publisher: SAGE Publications Ltd STM, ISSN: 0278-3649. DOI: 10.1177/0278364916688256. [Online]. Available: https://doi.org/10.1177/0278364916688256 (visited on 05/29/2020).

[127] H. Liu, Y. Zhang, W. Si, X. Xie, Y. Zhu, and S.-C. Zhu, "Interactive Robot Knowledge Patching Using Augmented Reality," in **2018 IEEE International Conference on Robotics and Automation (ICRA)**, ISSN: 2577-087X, May 2018, pp. 1947–1954. DOI: 10.1109/ICRA.2018.8462837.

[128] J. Scholtz, B. Antonishek, and J. Young, "Evaluation of a human-robot interface: Development of a situational awareness methodology," in **37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the**, Jan. 2004, 9 pp.–. DOI: 10.1109/HICSS.2004.1265327.

[129] J. Scholtz, B. Antonishek, and J. Young, "Implementation of a situation awareness assessment tool for evaluation of human-robot interfaces," **IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans**, vol. 35, no. 4, pp. 450–459, Jul. 2005, Conference Name: IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, ISSN: 1558-2426. DOI: 10.1109/TSMCA.2005.850589.

[130] R. Kirby, P. Rushton, C. Smith, **et al.**, **Wheelchair skills program manual version 5.1**, 2020. [Online]. Available: www.wheelchairskillsprogram.ca/eng/manual.php.

[131] C. P. Quintero, S. Li, M. K. Pan, W. P. Chan, H. Machiel Van der Loos, and E. Croft, "Robot programming through augmented trajectories in augmented reality," in **2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Oct. 2018, pp. 1838–1844. DOI: 10.1109/IROS.2018.8593700.

[132] H. Dinh, Q. Yuan, I. Vietcheslav, and G. Seet, "Augmented reality interface for taping robot," in **2017 18th International Conference on Advanced Robotics (ICAR)**, Jul. 2017, pp. 275–280. DOI: 10.1109/ICAR.2017.8023530.

[133] N. Palmer and E. Burchard, **Underrepresented Populations in Research**, en, 2020. [Online]. Available: https://recruit.ucsf.edu/underrepresented-populations-research (visited on 09/25/2020).

[134] T. Williams, D. Szafir, and T. Chakraborti, "The reality-virtuality interaction cube," in **Proceedings of the 2nd International Workshop on Virtual, Augmented, and Mixed Reality for HRI**, 2019.

[135] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, "Augmented reality: A class of displays on the reality-virtuality continuum," in **Telemanipulator and telepresence technologies**, International Society for Optics and Photonics, vol. 2351, 1995, pp. 282–292.

[136] "A Roadmap for US Robotics: From Internet to Robotics," Tech. Rep., Sep. 2020. [Online]. Available: `http://www.hichristensen.com/pdf/roadmap-2020.pdf`.

[137] E. Rosen, D. Whitney, E. Phillips, **et al.**, "Communicating and controlling robot arm motion intent through mixed-reality head-mounted displays," en, **The International Journal of Robotics Research**, vol. 38, no. 12-13, pp. 1513–1526, Oct. 2019, Publisher: SAGE Publications Ltd STM, ISSN: 0278-3649. DOI: `10.1177/0278364919842925`. [Online]. Available: `https://doi.org/10.1177/0278364919842925` (visited on 05/29/2020).

[138] E. Oyama and N. Shiroma, "Behavior Navigation System for Use in harsh environments," in **2011 IEEE International Symposium on Safety, Security, and Rescue Robotics**, ISSN: 2374-3247, Nov. 2011, pp. 272–277. DOI: `10.1109/SSRR.2011.6106799`.

[139] A. Hietanen, R. Pieters, M. Lanz, J. Latokartano, and J.-K. Kämäräinen, "AR-based interaction for human-robot collaborative manufacturing," en, **Robotics and Computer-Integrated Manufacturing**, vol. 63, p. 101 891, Jun. 2020, ISSN: 0736-5845. DOI: `10.1016/j.rcim.2019.101891`. [Online]. Available: `http://www.sciencedirect.com/science/article/pii/S0736584519307355` (visited on 05/15/2020).

[140] D. Szafir, B. Mutlu, and T. Fong, "Communicating Directionality in Flying Robots," in **2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)**, Mar. 2015, pp. 19–26.

[141] D. Szafir, "Mediating Human-Robot Interactions with Virtual, Augmented, and Mixed Reality," en, in **Virtual, Augmented and Mixed Reality. Applications and Case**

Studies, J. Y. Chen and G. Fragomeni, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2019, pp. 124–149, ISBN: 978-3-030-21565-1. DOI: 10.1007/978-3-030-21565-1_9.

[142] J. F. Fisac, A. Bajcsy, S. L. Herbert, **et al.**, "Probabilistically Safe Robot Planning with Confidence-Based Human Predictions," **arXiv:1806.00109 [cs]**, May 2018, arXiv: 1806.00109. [Online]. Available: http://arxiv.org/abs/1806.00109 (visited on 05/28/2019).

[143] J. P. McIntire, P. R. Havig, and E. E. Geiselman, "Stereoscopic 3D displays and human performance: A comprehensive review," en, **Displays**, vol. 35, no. 1, pp. 18–26, Jan. 2014, ISSN: 0141-9382. DOI: 10.1016/j.displa.2013.10.004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0141938213000929 (visited on 09/18/2021).

[144] D. Szafir and D. A. Szafir, "Connecting Human-Robot Interaction and Data Visualization," in **Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction**, ser. HRI '21, New York, NY, USA: Association for Computing Machinery, Mar. 2021, pp. 281–292, ISBN: 978-1-4503-8289-2. DOI: 10.1145/3434073.3444683. [Online]. Available: https://doi.org/10.1145/3434073.3444683 (visited on 09/16/2021).

[145] M. Colley, S. Krauss, M. Lanzer, and E. Rukzio, "How Should Automated Vehicles Communicate Critical Situations? A Comparative Analysis of Visualization Concepts," **Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies**, vol. 5, no. 3, 94:1–94:23, Sep. 2021. DOI: 10.1145/3478111. [Online]. Available: https://doi.org/10.1145/3478111 (visited on 09/18/2021).

[146] E. Rosen, D. Whitney, M. Fishman, D. Ullman, and S. Tellex, "Mixed Reality as a Bidirectional Communication Interface for Human-Robot Interaction," in **2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, ISSN: 2153-0866, Oct. 2020, pp. 11 431–11 438. DOI: 10.1109/IROS45743.2020.9340822.

[147] K. Chandan, V. Kudalkar, X. Li, and S. Zhang, "Negotiation-based Human-Robot Collaboration via Augmented Reality," **arXiv:1909.11227 [cs]**, Mar. 2020, arXiv: 1909.11227. [Online]. Available: `http://arxiv.org/abs/1909.11227` (visited on 09/18/2021).

[148] T. A. Sitompul and M. Wallmyr, "Using augmented reality to improve productivity and safety for heavy machinery operators: State of the art," in **The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry**, ser. VRCAI '19, New York, NY, USA: Association for Computing Machinery, Nov. 14, 2019, pp. 1–9, ISBN: 978-1-4503-7002-8. DOI: `10.1145/3359997.3365689`. [Online]. Available: `https://doi.org/10.1145/3359997.3365689` (visited on 09/18/2021).

[149] X. Li, W. Yi, H.-L. Chi, X. Wang, and A. P. C. Chan, "A critical review of virtual and augmented reality (VR/AR) applications in construction safety," en, **Automation in Construction**, vol. 86, pp. 150–162, Feb. 2018, ISSN: 0926-5805. DOI: `10.1016/j.autcon.2017.11.003`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0926580517309962` (visited on 09/18/2021).

[150] D. Tatić and B. Tešić, "The application of augmented reality technologies for the improvement of occupational safety in an industrial environment," en, **Computers in Industry**, vol. 85, pp. 1–10, Feb. 2017, ISSN: 0166-3615. DOI: `10.1016/j.compind.2016.11.004`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/S0166361516302718` (visited on 09/18/2021).

[151] **Hard Hat for HoloLens 2 Solution**, en-US, Running Time: 23. [Online]. Available: `https://visuallive.com/hard-hat-for-hololens-2-system/` (visited on 09/29/2021).

[152] **Trimble XR10 with HoloLens 2**, en-us. [Online]. Available: `https://www.microsoft.com/en-us/d/trimble-xr10-with-hololens-2/8smjj5mx7zt7` (visited on 09/29/2021).

[153] S. H. Choi, K.-B. Park, D. H. Roh, **et al.**, "An integrated mixed reality system for safety-aware human-robot collaboration using deep learning and digital twin generation," **Robotics and Computer-Integrated Manufacturing**, vol. 73, p. 102 258, Feb. 1, 2022, ISSN: 0736-5845.

DOI: 10.1016/j.rcim.2021.102258. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0736584521001381 (visited on 09/20/2021).

[154] N. Scheiber, "Inside an amazon warehouse, robots' ways rub off on humans," **The New York Times**, Jul. 3, 2019, ISSN: 0362-4331. [Online]. Available: https://www.nytimes.com/2019/07/03/business/economy/amazon-warehouse-labor-robots.html (visited on 09/18/2021).

[155] A. Delfanti and B. Frey, "Humanly extended automation or the future of work seen through amazon patents," **Science, Technology, & Human Values**, vol. 46, no. 3, pp. 655–682, May 1, 2021, Publisher: SAGE Publications Inc, ISSN: 0162-2439. DOI: 10.1177/0162243920943665. [Online]. Available: https://doi.org/10.1177/0162243920943665 (visited on 09/18/2021).

[156] U. Madan, M. E. Bundy, D. D. Glick, and J. E. Darrow, "Augmented reality user interface facilitating fulfillment," U.S. Patent 10,055,645 B1, Aug. 21, 2018.

[157] U. Technologies, **Unity Real-Time Development Platform — 3D, 2D VR & AR Engine**, en. [Online]. Available: https://unity.com/ (visited on 09/29/2021).

[158] C. Fairhurst, C. E. Hewitt, and D. J. Torgerson, "Using pairwise randomisation to reduce the risk of bias," **Research Methods in Medicine & Health Sciences**, vol. 1, no. 1, pp. 2–6, Sep. 2020, Publisher: SAGE Publications Ltd STM, ISSN: 2632-0843. DOI: 10.1177/2632084319884178. [Online]. Available: https://doi.org/10.1177/2632084319884178 (visited on 09/26/2021).

[159] L. Veling and C. McGinn, "Qualitative Research in HRI: A Review and Taxonomy," en, **International Journal of Social Robotics**, Feb. 2021, ISSN: 1875-4805. DOI: 10.1007/s12369-020-00723-z. [Online]. Available: https://doi.org/10.1007/s12369-020-00723-z (visited on 09/29/2021).

[160] M. Luria, J. Forlizzi, and J. Hodgins, "The Effects of Eye Design on the Perception of Social Robots," in **2018 27th IEEE International Symposium on Robot and Human**

Interactive Communication (RO-MAN), ISSN: 1944-9437, Aug. 2018, pp. 1032–1037. DOI: 10.1109/ROMAN.2018.8525767.

[161] K. S. Welfare, M. R. Hallowell, J. A. Shah, and L. D. Riek, "Consider the Human Work Experience When Integrating Robotics in the Workplace," in **2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)**, ISSN: 2167-2148, Mar. 2019, pp. 75–84. DOI: 10.1109/HRI.2019.8673139.

[162] D. Silvera-Tawil, D. Bradford, and C. Roberts-Yates, "Talk to Me: The Role of Human-Robot Interaction in Improving Verbal Communication Skills in Students with Autism or Intellectual Disability," in **2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)**, ISSN: 1944-9437, Aug. 2018, pp. 1–6. DOI: 10.1109/ROMAN.2018.8525698.

[163] Ö. Örün and Y. Akbulut, "Effect of multitasking, physical environment and electroencephalography use on cognitive load and retention," en, **Computers in Human Behavior**, vol. 92, pp. 216–229, Mar. 2019, ISSN: 0747-5632. DOI: 10.1016/j.chb.2018.11.027. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563218305624 (visited on 11/11/2021).

[164] P. Robinette, W. Li, R. Allen, A. M. Howard, and A. R. Wagner, "Overtrust of robots in emergency evacuation scenarios," in **2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)**, ISSN: 2167-2148, Mar. 2016, pp. 101–108. DOI: 10.1109/HRI.2016.7451740.

[165] J. D. Lee and K. A. See, "Trust in Automation: Designing for Appropriate Reliance," en, **Human Factors**, vol. 46, no. 1, pp. 50–80, Mar. 2004, Publisher: SAGE Publications Inc, ISSN: 0018-7208. DOI: 10.1518/hfes.46.1.50_30392. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1518/hfes.46.1.5030392 (visited on 09/27/2021).

[166] N. Pidgeon, J. Walls, A. Weyman, and T. Horlick-Jones, **Perceptions of and Trust in the Health and Safety Executive as a Risk Regulator**. Health and Safety Executive, 2003.

[167] B. C. Gunia, S. H. Kim, and K. M. Sutcliffe, “Trust in Safety-Critical Contexts,” in **The Routledge Companion to Trust**, Num Pages: 15, Routledge, 2018, ISBN: 978-1-315-74557-2.

[168] B. Hayes and M. Moniz, “Trustworthy Human-Centered Automation Through Explainable AI and High-Fidelity Simulation,” en, in **Advances in Simulation and Digital Human Modeling**, D. N. Cassenti, S. Scataglini, S. L. Rajulu, and J. L. Wright, Eds., ser. Advances in Intelligent Systems and Computing, Cham: Springer International Publishing, 2021, pp. 3–9, ISBN: 978-3-030-51064-0. DOI: 10.1007/978-3-030-51064-0_1.

[169] “AI legal research tools — lexis+ — LexisNexis.” (), [Online]. Available: https://www.lexisnexis.com/en-us/products/lexis-plus.page (visited on 10/30/2023).

[170] “Artificial intelligence (AI) services & solutions — accenture.” (), [Online]. Available: https://www.accenture.com/us-en/services/ai-artificial-intelligence-index (visited on 10/30/2023).

[171] “IBM watsonx - an AI and data platform built for business.” (), [Online]. Available: https://www.ibm.com/watsonx (visited on 10/30/2023).

[172] “Healthcare.ai.” (), [Online]. Available: https://healthcare.ai/ (visited on 10/30/2023).

[173] S. Bhasker, D. Bruce, J. Lamb, and G. Stein, **Tackling healthcare’s biggest burdens with generative ai**, Jul. 2023. [Online]. Available: https://www.mckinsey.com/industries/healthcare/our-insights/tackling-healthcares-biggest-burdens-with-generative-ai.

[174] A. Schroer, **33 examples of ai in finance**, Feb. 2024. [Online]. Available: https://builtin.com/artificial-intelligence/ai-finance-banking-applications-companies.

[175] National Association of Realtors, **Artificial intelligence (AI) in real estate**, www.nar.realtor, Jun. 28, 2023. [Online]. Available: https://www.nar.realtor/artificial-intelligence-real-estate (visited on 10/30/2023).

[176] Accenture, **Unleash the power of data with ai.retail — accenture**, 2023. [Online]. Available: `https://www.accenture.com/us-en/services/retail/applied-intelligence-retail` (visited on 10/30/2023).

[177] Intel, **Artificial intelligence (AI) in retail**, Intel, 2023. [Online]. Available: `https://www.intel.com/content/www/us/en/retail/solutions/ai-in-retail.html` (visited on 10/30/2023).

[178] D. Marotta, **Artificial intelligence: How AI is changing retail**, Hitachi Solutions, Jun. 10, 2020. [Online]. Available: `https://global.hitachi-solutions.com/blog/ai-in-retail/` (visited on 10/30/2023).

[179] Department of Education, **Artificial intelligence**, Office of Educational Technology, 2023. [Online]. Available: `https://tech.ed.gov/ai/` (visited on 10/30/2023).

[180] R. Emsley, "ChatGPT: These are not hallucinations – they're fabrications and falsifications," en, **Schizophrenia**, vol. 9, no. 1, pp. 1–2, Aug. 2023, Number: 1 Publisher: Nature Publishing Group, ISSN: 2754-6993. DOI: `10.1038/s41537-023-00379-4`. [Online]. Available: `https://www.nature.com/articles/s41537-023-00379-4` (visited on 11/07/2023).

[181] M. Phuong, M. Aitchison, E. Catt, **et al.**, **Evaluating Frontier Models for Dangerous Capabilities**, arXiv:2403.13793 [cs], Mar. 2024. [Online]. Available: `http://arxiv.org/abs/2403.13793` (visited on 03/25/2024).

[182] Y. Ye, H. You, and J. Du, "Improved trust in human-robot collaboration with ChatGPT," **IEEE Access**, vol. 11, pp. 55 748–55 754, 2023, ISSN: 2169-3536. DOI: `10.1109/ACCESS.2023.3282111`. [Online]. Available: `https://ieeexplore.ieee.org/document/10141597/` (visited on 11/17/2023).

[183] A. Brohan, N. Brown, J. Carbajal, **et al.**, "RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control," en, Jul. 2023. [Online]. Available: `https://robotics-transformer2.github.io/`.

[184] K. Rana, J. Haviland, S. Garg, J. Abou-Chakra, I. Reid, and N. Suenderhauf, "SayPlan: Grounding Large Language Models using 3D Scene Graphs for Scalable Robot Task Planning," en, Aug. 2023. [Online]. Available: https://openreview.net/forum?id=wMpOMO0Ss7a (visited on 11/06/2023).

[185] A. Szot, M. Schwarzer, H. Agrawal, **et al.**, **Large Language Models as Generalizable Policies for Embodied Tasks**, arXiv:2310.17722 [cs], Oct. 2023. [Online]. Available: http://arxiv.org/abs/2310.17722 (visited on 11/06/2023).

[186] M. Gombolay, X. J. Yang, B. Hayes, **et al.**, "Robotic assistance in the coordination of patient care," en, **The International Journal of Robotics Research**, vol. 37, no. 10, pp. 1300–1316, Sep. 2018, Publisher: SAGE Publications Ltd STM, ISSN: 0278-3649. DOI: 10.1177/0278364918778344. [Online]. Available: https://doi.org/10.1177/0278364918778344 (visited on 10/19/2023).

[187] J. Zamfirescu-Pereira, R. Y. Wong, B. Hartmann, and Q. Yang, "Why johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts," in **Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems**, ser. CHI '23, New York, NY, USA: Association for Computing Machinery, Apr. 19, 2023, pp. 1–21, ISBN: 978-1-4503-9421-5. DOI: 10.1145/3544548.3581388. [Online]. Available: https://dl.acm.org/doi/10.1145/3544548.3581388 (visited on 08/23/2023).

[188] F. Dell'Acqua, E. McFowland, E. R. Mollick, **et al.**, **Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality**, en, SSRN Scholarly Paper, Rochester, NY, Sep. 2023. DOI: 10.2139/ssrn.4573321. [Online]. Available: https://papers.ssrn.com/abstract=4573321 (visited on 10/09/2023).

[189] S. Giannini, "Generative AI and the future of education," UNESCO, Tech. Rep., Jul. 2023. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000385877 (visited on 10/30/2023).

[190] T. A. Bach, A. Khan, H. Hallock, G. Beltrão, and S. Sousa, “A Systematic Literature Review of User Trust in AI-Enabled Systems: An HCI Perspective,” **International Journal of Human–Computer Interaction**, vol. 40, no. 5, pp. 1251–1266, Mar. 2024, Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2022.2138826, ISSN: 1044-7318. DOI: 10.1080/10447318.2022.2138826. [Online]. Available: https://doi.org/10.1080/10447318.2022.2138826 (visited on 03/24/2024).

[191] T. Chakraborti, A. Kulkarni, S. Sreedharan, D. E. Smith, and S. Kambhampati, “Explicability? Legibility? Predictability? Transparency? Privacy? Security? The Emerging Landscape of Interpretable Agent Behavior,” en, **Proceedings of the International Conference on Automated Planning and Scheduling**, vol. 29, pp. 86–96, 2019, ISSN: 2334-0843. DOI: 10.1609/icaps.v29i1.3463. [Online]. Available: https://ojs.aaai.org/index.php/ICAPS/article/view/3463 (visited on 10/20/2022).

[192] H. Vasconcelos, M. Jörke, M. Grunde-McLaughlin, T. Gerstenberg, M. S. Bernstein, and R. Krishna, “Explanations Can Reduce Overreliance on AI Systems During Decision-Making,” **Proceedings of the ACM on Human-Computer Interaction**, vol. 7, no. CSCW1, 129:1–129:38, Apr. 2023. DOI: 10.1145/3579605. [Online]. Available: https://dl.acm.org/doi/10.1145/3579605 (visited on 11/20/2023).

[193] S. Mehrotra, C. C. Jorge, C. M. Jonker, and M. L. Tielman, “Integrity-based Explanations for Fostering Appropriate Trust in AI Agents,” en, **ACM Transactions on Interactive Intelligent Systems**, vol. 14, no. 1, pp. 1–36, Mar. 2024, ISSN: 2160-6455, 2160-6463. DOI: 10.1145/3610578. [Online]. Available: https://dl.acm.org/doi/10.1145/3610578 (visited on 03/25/2024).

[194] M. Luria, “Co-design perspectives on algorithm transparency reporting: Guidelines and prototypes,” in **Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency**, ser. FAccT ’23, New York, NY, USA: Association for Computing Ma-

chinery, Jun. 12, 2023, pp. 1076–1087. DOI: 10.1145/3593013.3594064. [Online]. Available: https://dl.acm.org/doi/10.1145/3593013.3594064 (visited on 10/05/2023).

[195] D. Dunning, C. Heath, and J. M. Suls, "Flawed Self-Assessment: Implications for Health, Education, and the Workplace," en, **Psychological Science in the Public Interest**, vol. 5, no. 3, pp. 69–106, Dec. 2004, Publisher: SAGE Publications Inc, ISSN: 1529-1006. DOI: 10.1111/j.1529-1006.2004.00018.x. [Online]. Available: https://doi.org/10.1111/j.1529-1006.2004.00018.x (visited on 03/25/2024).

[196] J. Ehrlinger, K. Johnson, M. Banner, D. Dunning, and J. Kruger, "Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent," **Organizational Behavior and Human Decision Processes**, vol. 105, no. 1, pp. 98–121, Jan. 2008, ISSN: 0749-5978. DOI: 10.1016/j.obhdp.2007.05.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S074959780700060X (visited on 03/25/2024).

[197] D. Dunning, "Chapter five - The Dunning–Kruger Effect: On Being Ignorant of One's Own Ignorance," in **Advances in Experimental Social Psychology**, J. M. Olson and M. P. Zanna, Eds., vol. 44, Academic Press, Jan. 2011, pp. 247–296. DOI: 10.1016/B978-0-12-385522-0.00005-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780123855220000056 (visited on 03/25/2024).

[198] L. Konstantinou, D. Panos, and E. Karapanos, "Exploring the Design of Technology-Mediated Nudges for Online Misinformation," **International Journal of Human–Computer Interaction**, vol. 0, no. 0, pp. 1–28, 2024, Publisher: Taylor & Francis eprint, ISSN: 1044-7318. DOI: 10.1080/10447318.2023.2301265. [Online]. Available: https://doi.org/10.1080/10447318.2023.2301265 (visited on 02/01/2024).

[199] M. Weinmann, C. Schneider, and J. v. Brocke, "Digital Nudging," en, **Business & Information Systems Engineering**, vol. 58, no. 6, pp. 433–436, Dec. 2016, ISSN: 1867-0202. DOI:

10.1007/s12599-016-0453-1. [Online]. Available: https://doi.org/10.1007/s12599-016-0453-1 (visited on 03/25/2024).

[200] A. Caraban, E. Karapanos, D. Gonçalves, and P. Campos, "23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction," en, in **Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems**, Glasgow Scotland Uk: ACM, May 2019, pp. 1–15, ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300733. [Online]. Available: https://dl.acm.org/doi/10.1145/3290605.3300733 (visited on 03/25/2024).

[201] P. B. Brandtzaeg and A. Følstad, "Why People Use Chatbots," en, in **Internet Science**, I. Kompatsiaris, J. Cave, A. Satsiou, **et al.**, Eds., Cham: Springer International Publishing, 2017, pp. 377–392, ISBN: 978-3-319-70284-1. DOI: 10.1007/978-3-319-70284-1_30.

[202] N. C. Benda, L. L. Novak, C. Reale, and J. S. Ancker, "Trust in AI: Why we should be designing for APPROPRIATE reliance," **Journal of the American Medical Informatics Association**, vol. 29, no. 1, pp. 207–212, Jan. 2022, ISSN: 1527-974X. DOI: 10.1093/jamia/ocab238. [Online]. Available: https://doi.org/10.1093/jamia/ocab238 (visited on 03/25/2024).

[203] K. Okamura and S. Yamada, "Adaptive trust calibration for human-AI collaboration," en, **PLOS ONE**, vol. 15, no. 2, C. Lv, Ed., e0229132, Feb. 2020, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0229132. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0229132 (visited on 03/25/2024).

[204] K. Okamura and S. Yamada, "Calibrating Trust in Human-Drone Cooperative Navigation," en, in **2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)**, Naples, Italy: IEEE, Aug. 2020, pp. 1274–1279, ISBN: 978-1-72816-075-7. DOI: 10.1109/RO-MAN47096.2020.9223509. [Online]. Available: https://ieeexplore.ieee.org/document/9223509/ (visited on 03/25/2024).

[205] A. L. Cox, S. J. Gould, M. E. Cecchinato, I. Iacovides, and I. Renfree, "Design Frictions for Mindful Interactions: The Case for Microboundaries," en, in **Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems**, San Jose California USA: ACM, May 2016, pp. 1389–1397, ISBN: 978-1-4503-4082-3. DOI: 10.1145/2851581.2892410. [Online]. Available: https://dl.acm.org/doi/10.1145/2851581.2892410 (visited on 12/13/2023).

[206] C. Guo, N. Zheng, and C. (Guo, "Seeing is Not Believing: A Nuanced View of Misinformation Warning Efficacy on Video-Sharing Social Media Platforms," en, **Proceedings of the ACM on Human-Computer Interaction**, vol. 7, no. CSCW2, pp. 1–35, Sep. 2023, ISSN: 2573-0142. DOI: 10.1145/3610085. [Online]. Available: https://dl.acm.org/doi/10.1145/3610085 (visited on 12/13/2023).

[207] P. Mena, "Cleaning Up Social Media: The Effect of Warning Labels on Likelihood of Sharing False News on Facebook," en, **Policy & Internet**, vol. 12, no. 2, pp. 165–183, 2020, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/poi3.214, ISSN: 1944-2866. DOI: 10.1002/poi3.214. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/poi3.214 (visited on 03/25/2024).

[208] A. Vance, J. L. Jenkins, B. B. Anderson, D. K. Bjornn, and C. B. Kirwan, "Tuning Out Security Warnings: A Longitudinal Examination of Habituation Through fMRI, Eye Tracking, and Field Experiments," en, **MIS Quarterly**, vol. 42, no. 2, pp. 355–380, Feb. 2018, ISSN: 02767783, 21629730. DOI: 10.25300/MISQ/2018/14124. [Online]. Available: https://misq.org/skin/frontend/default/misq/pdf/appendices/2018/V42I2Appendices/01_14124_RA_VanceJenkins.pdf (visited on 12/13/2023).

[209] Z. Buçinca, M. B. Malaya, and K. Z. Gajos, "To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making," **Proceedings of the ACM on Human-Computer Interaction**, vol. 5, no. CSCW1, 188:1–188:21, Apr. 2021.

DOI: 10.1145/3449287. [Online]. Available: https://dl.acm.org/doi/10.1145/3449287 (visited on 11/20/2023).

[210] K. Parsakia, "The Effect of Chatbots and AI on The Self-Efficacy, Self-Esteem, Problem-Solving and Critical Thinking of Students," en, **Health Nexus**, vol. 1, no. 1, pp. 71–76, Jan. 2023, Number: 1, ISSN: 2981-2569. DOI: 10.61838/hn.1.1.14. [Online]. Available: https://journals.kmanpub.com/index.php/Health-Nexus/article/view/908 (visited on 03/28/2024).

[211] A. Rapp, L. Curti, and A. Boldi, "The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots," **International Journal of Human-Computer Studies**, vol. 151, p. 102 630, Jul. 2021, ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2021.102630. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1071581921000483 (visited on 03/28/2024).

[212] W. M. Lim, A. Gunasekara, J. L. Pallant, J. I. Pallant, and E. Pechenkina, "Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators," **The International Journal of Management Education**, vol. 21, no. 2, p. 100 790, Jul. 2023, ISSN: 1472-8117. DOI: 10.1016/j.ijme.2023.100790. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1472811723000289 (visited on 03/25/2024).

[213] D. Ullman and B. F. Malle, "Measuring Gains and Losses in Human-Robot Trust: Evidence for Differentiable Components of Trust," in **2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)**, ISSN: 2167-2148, Mar. 2019, pp. 618–619. DOI: 10.1109/HRI.2019.8673154. [Online]. Available: https://ieeexplore.ieee.org/document/8673154 (visited on 02/24/2024).

[214] V. Braun and V. Clarke, "Using thematic analysis in psychology," **Qualitative Research in Psychology**, vol. 3, no. 2, pp. 77–101, Jan. 2006, ISSN: 1478-0887. DOI: 10.1191/

1478088706qp063oa. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa (visited on 10/13/2023).

[215] J. S. Park, R. Barber, A. Kirlik, and K. Karahalios, "A Slow Algorithm Improves Users' Assessments of the Algorithm's Accuracy," en, **Proceedings of the ACM on Human-Computer Interaction**, vol. 3, no. CSCW, pp. 1–15, Nov. 2019, ISSN: 2573-0142. DOI: 10.1145/3359204. [Online]. Available: https://dl.acm.org/doi/10.1145/3359204 (visited on 12/13/2023).

[216] J. L. Nelson and S. C. Lewis, "Only "sheep" trust journalists? how citizens' self-perceptions shape their approach to news," **New Media & Society**, vol. 25, no. 7, pp. 1522–1541, Jul. 1, 2023, Publisher: SAGE Publications, ISSN: 1461-4448. DOI: 10.1177/14614448211018160. [Online]. Available: https://doi.org/10.1177/14614448211018160 (visited on 03/18/2024).

[217] N. Earth Science Data Systems, **FIRMS Frequently Asked Questions — Earthdata**, en, Basic Page, Publisher: Earth Science Data Systems, NASA, Dec. 2021. [Online]. Available: https://www.earthdata.nasa.gov/faq/firms-faq (visited on 10/17/2023).

[218] **California Forest Observatory**, en, Aug. 2019. [Online]. Available: https://salo.ai/projects/california-forest-observatory (visited on 10/17/2023).

[219] **Pano AI**. [Online]. Available: https://www.pano.ai/ (visited on 10/17/2023).

[220] M. Kalakrishnan, S. Chitta, E. Theodorou, P. Pastor, and S. Schaal, "STOMP: Stochastic trajectory optimization for motion planning," in **2011 IEEE International Conference on Robotics and Automation**, ISSN: 1050-4729, May 2011, pp. 4569–4574. DOI: 10.1109/ICRA.2011.5980280.

[221] H. M. Ray, R. Singer, and N. Ahmed, "A review of the operational use of UAS in public safety emergency incidents," in **2022 International Conference on Unmanned Aircraft Systems (ICUAS)**, ISSN: 2575-7296, Jun. 2022, pp. 922–931. DOI: 10.1109/ICUAS54217.2022.9836061. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9836061 (visited on 11/16/2023).

[222] J. Reinhardt, A. Pereira, D. Beckert, and K. Bengler, "Dominance and movement cues of robot motion: A user study on trust and predictability," in **2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)**, Banff, AB: IEEE, Oct. 2017, pp. 1493–1498, ISBN: 978-1-5386-1645-1. DOI: 10.1109/SMC.2017.8122825. [Online]. Available: http://ieeexplore.ieee.org/document/8122825/ (visited on 11/15/2023).

[223] M. Chandarana, E. L. Meszaros, A. Trujillo, and B. D. Allen, ""Fly Like This": Natural Language Interfaces for UAV Mission Planning," en, Nice, France, Mar. 2017. [Online]. Available: https://ntrs.nasa.gov/citations/20170002593.

[224] GeoNadir, **How to use DroneLink for a drone mapping mission**, Aug. 2022. [Online]. Available: https://www.youtube.com/watch?v=bV7E2Q5Fi4U (visited on 06/11/2024).

[225] S. Thellman and T. Ziemke, "The perceptual belief problem: Why explainability is a tough challenge in social robotics," **ACM Transactions on Human-Robot Interaction (THRI)**, vol. 10, no. 3, pp. 1–15, 2021.

[226] S. Sjøberg, "Constructivism and learning," **International encyclopedia of education**, vol. 5, pp. 485–490, 2010.

[227] L. H. Lewis and C. J. Williams, "Experiential learning: Past and present," **New directions for adult and continuing education**, vol. 1994, no. 62, pp. 5–16, 1994.

[228] D. Wood, J. S. Bruner, and G. Ross, "The role of tutoring in problem solving*," **Journal of Child Psychology and Psychiatry**, vol. 17, no. 2, pp. 89–100, 1976, ISSN: 1469-7610. DOI: 10.1111/j.1469-7610.1976.tb00381.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-7610.1976.tb00381.x (visited on 06/06/2024).

[229] E. A. Davis, "Scaffolding learning," in **Science Teacher Education in Mainland China**, Jan. 1, 2015, pp. 845–847. DOI: 10.1007/978-94-007-2150-0_531.

[230] N. F. Jumaat and Z. Tasir, "Instructional scaffolding in online learning environment: A meta-analysis," in **2014 International Conference on Teaching and Learning in**

Computing and Engineering, Apr. 2014, pp. 74–77. DOI: 10.1109/LaTiCE.2014.22. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6821832 (visited on 06/06/2024).

[231] B. Morrell, R. Thakker, G. Merewether, **et al.**, "Comparison of trajectory optimization algorithms for high-speed quadrotor flight near obstacles," **IEEE Robotics and Automation Letters**, vol. 3, no. 4, pp. 4399–4406, Oct. 2018, Conference Name: IEEE Robotics and Automation Letters, ISSN: 2377-3766. DOI: 10.1109/LRA.2018.2868866. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8454815 (visited on 05/26/2024).

[232] P. Foehn, A. Romero, and D. Scaramuzza, "Time-optimal planning for quadrotor waypoint flight," **Science Robotics**, vol. 6, no. 56, eabh1221, Jul. 21, 2021, Publisher: American Association for the Advancement of Science. DOI: 10.1126/scirobotics.abh1221. [Online]. Available: https://www.science.org/doi/full/10.1126/scirobotics.abh1221 (visited on 05/23/2024).

[233] G. E. Chamitoff, A. Saenz-Otero, J. G. Katz, S. Ulrich, B. J. Morrell, and P. W. Gibbens, "Real-time maneuver optimization of space-based robots in a dynamic environment: Theory and on-orbit experiments," **Acta Astronautica**, vol. 142, pp. 170–183, Jan. 1, 2018, ISSN: 0094-5765. DOI: 10.1016/j.actaastro.2017.10.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0094576516300431 (visited on 05/26/2024).

[234] T. Ma, H. Zhou, B. Qian, and A. Fu, "A large-scale clustering and 3d trajectory optimization approach for UAV swarms," **Science China Information Sciences**, vol. 64, no. 4, p. 140 306, Mar. 5, 2021, ISSN: 1869-1919. DOI: 10.1007/s11432-020-3013-1. [Online]. Available: https://doi.org/10.1007/s11432-020-3013-1 (visited on 05/26/2024).

[235] P. Pradeep, S. G. Park, and P. Wei, "Trajectory optimization of multirotor agricultural UAVs," in **2018 IEEE Aerospace Conference**, Big Sky, MT: IEEE, Mar. 2018, pp. 1–7, ISBN: 978-1-5386-2014-4. DOI: 10.1109/AERO.2018.8396617. [Online]. Available: https://ieeexplore.ieee.org/document/8396617/ (visited on 05/23/2024).

[236] A. Brown and D. Anderson, "Trajectory optimization for high-altitude long-endurance uav maritime radar surveillance," **IEEE Transactions on Aerospace and Electronic Systems**, vol. 56, no. 3, pp. 2406–2421, 2020. DOI: 10.1109/TAES.2019.2949384.

[237] G. Bevacqua, J. Cacace, A. Finzi, and V. Lippiello, "Mixed-initiative planning and execution for multiple drones in search and rescue missions," **Proceedings of the International Conference on Automated Planning and Scheduling**, vol. 25, pp. 315–323, Apr. 8, 2015, ISSN: 2334-0843. DOI: 10.1609/icaps.v25i1.13700. [Online]. Available: https://ojs.aaai.org/index.php/ICAPS/article/view/13700 (visited on 05/23/2024).

[238] H. M. Ray, Z. Laouar, Z. Sunberg, and N. Ahmed, **Human-centered autonomy for UAS target search**, Mar. 6, 2024. DOI: 10.48550/arXiv.2309.06395. arXiv: 2309.06395[cs]. [Online]. Available: http://arxiv.org/abs/2309.06395 (visited on 05/23/2024).

[239] H. Lee and S. Park, "Sensing-aware deep reinforcement learning with HCI-based human-in-the-loop feedback for autonomous nonlinear drone mobility control," **IEEE Access**, vol. 12, pp. 1727–1736, Jan. 1, 2024. DOI: 10.1109/ACCESS.2023.3346917.

[240] D. Gopinath, S. Jain, and B. D. Argall, "Human-in-the-loop optimization of shared autonomy in assistive robotics," **IEEE Robotics and Automation Letters**, vol. 2, no. 1, pp. 247–254, Jan. 2017, Conference Name: IEEE Robotics and Automation Letters, ISSN: 2377-3766. DOI: 10.1109/LRA.2016.2593928. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7518989 (visited on 05/26/2024).

[241] D. F. N. Gordon, C. McGreavy, A. Christou, and S. Vijayakumar, "Human-in-the-loop optimization of exoskeleton assistance via online simulation of metabolic cost," **IEEE Transactions on Robotics**, vol. 38, no. 3, pp. 1410–1429, Jun. 2022, Conference Name: IEEE Transactions on Robotics, ISSN: 1941-0468. DOI: 10.1109/TRO.2021.3133137. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9698243 (visited on 05/26/2024).

[242] Y. Zhou, Y. Zhang, X. Luo, and M. M. Zavlanos, "Human-in-the-loop robot planning with non-contextual bandit feedback," in **2021 60th IEEE Conference on Decision and Control (CDC)**, ISSN: 2576-2370, Dec. 2021, pp. 2848–2853. DOI: `10.1109/CDC45484.2021.9683023`. [Online]. Available: `https://ieeexplore.ieee.org/abstract/document/9683023` (visited on 05/26/2024).

[243] T. M. Howard, S. Tellex, and N. Roy, "A natural language planner interface for mobile manipulators," in **2014 IEEE International Conference on Robotics and Automation (ICRA)**, ISSN: 1050-4729, May 2014, pp. 6652–6659. DOI: `10.1109/ICRA.2014.6907841`. [Online]. Available: `https://ieeexplore.ieee.org/abstract/document/6907841` (visited on 11/29/2023).

[244] W. Yu, N. Gileadi, C. Fu, **et al.**, **Language to Rewards for Robotic Skill Synthesis**, arXiv:2306.08647 [cs], Jun. 2023. DOI: `10.48550/arXiv.2306.08647`. [Online]. Available: `http://arxiv.org/abs/2306.08647` (visited on 11/07/2023).

[245] **Artificial intelligence facial recognition**, ed. by C. Hansen, J. Buckner, J. Bacon, and K. Tipper.

[246] C. T. Chang and B. Hayes, "Safety and accountability for large language model use in HRI," 2024.