ASTRID: A Robotic Tutor for Nurse Training to Reduce Healthcare-Associated Infections

Peizhu Qian[*], Filip Bajraktari[*], Carlos Quintero-Peña[*], Qingxi Meng[*],
Shannan Hamlin[†], Lydia Kavraki[‡], and Vaibhav Unhelkar[‡]
[*]Department of Computer Science, Rice University, Houston, Texas USA
pqian@rice.edu, filip.bajraktari@rice.edu, carlosq@rice.edu, qm15@rice.edu
[†]Center for Nursing Research, Education and Practice, Houston Methodist, Houston, Texas USA
shamlin@houstonmethodist.org
[‡]Ken Kennedy Institute, Rice University, Houston, Texas USA
kavraki@rice.edu, unhelkar@rice.edu

***Abstract*—The central line dressing change is a life-critical procedure performed by nurses to provide patients with rapid infusion of fluids, such as blood and medications. Due to their complexity and the heavy workloads nurses face, dressing changes are prone to preventable errors that can result in central line-associated bloodstream infections (CLABSIs), leading to serious health complications or, in the worst cases, patient death. In the post-COVID-19 era, CLABSI rates have increased, partly due to the heightened nursing workload caused by shortages of both registered nurses and nurse educators. To address this challenge, healthcare facilities are seeking innovative nurse training solutions to complement expert nurse educators.**

In response, we present the design, development and evaluation of a robotic tutoring system, ASTRID: the Automated Sterile Technique Review and Instruction Device. ASTRID, which is the outcome of a two-year participatory design process, is designed to aid in the training of nursing skills essential for CLABSI prevention. First, we describe insights gained from interviews with nurse educators and nurses, which revealed the gaps of current training methods and requirements for new training tools. Based on these findings, we outline the development of our robotic tutor, which interacts with nursing students, providing real-time interventions and summary feedback to support skill acquisition. Finally, we present evaluations of the system's performance and perceived usefulness, conducted in a simulated clinical setting with nurse participants. These evaluations demonstrate the potential of our robotic tutor in nursing education. Our work highlights the importance of participatory design for robotics systems, and motivates new avenues for foundational research in robotics.

***Index Terms*—Robotics in Healthcare, Human-Centered Robotics, Intelligent Tutors, Participatory Design**

Fig. 1. Artificial rendering of the training environment and ASTRID, which is composed of the Stretch robot, a depth camera, and a computer screen.

I. Introduction

As the largest sector of healthcare, nurses play a critical role in delivering high-quality patient care and maintaining the stability of the entire healthcare system. However, this stability is now at risk due to a significant global shortage of nursing professionals [90, 115, 54]. Among other challenges to patient health outcomes, this shortage is placing an increased burden on nursing educators to train a new generation of professionals, while hospitals are tasked with recruiting and onboarding a substantial number of new nurses each year.

Beyond formal education in nursing schools, hospitals allocate substantial resources to train new nurses in hospital-specific practices [66]. With the increasing complexity of patient care, nursing staff must be regularly upskilled. For example, the Houston Methodist hospital system trains over a thousand nurses each year. This process includes instruction from nursing educators, followed by skill refinement under the supervision of an experienced nurse mentor. Sustaining the traditional nurse-to-nurse training model is increasingly challenging [47, 101, 63].

As a result, healthcare facilities are actively exploring innovative solutions to enhance and support nursing education [107, 37, 38, 1, 42, 113]. We posit that *robotic tutors can help address this urgent need for nursing education*. Specifically, robotic tutors can allow nursing students to practice critical skills when expert nurses are unavailable, complementing nurse-to-nurse training. Examining this hypothesis, we present ASTRID: a robotic tutor for nursing education.

ASTRID is designed to help nurses acquire necessary skills to reduce the chances of healthcare-associated infections [62]. Healthcare-associated infections, which refer to infections that occur while the patient is receiving care, can lead to serious complications including death [130, 140, 26, 48]. The rates of these infections have exacerbated post-COVID-19, in part due to the nursing shortage. These infections, however, are largely preventable through meticulous care, adherence to protocols, and regular training — currently delivered through a nurse-to-

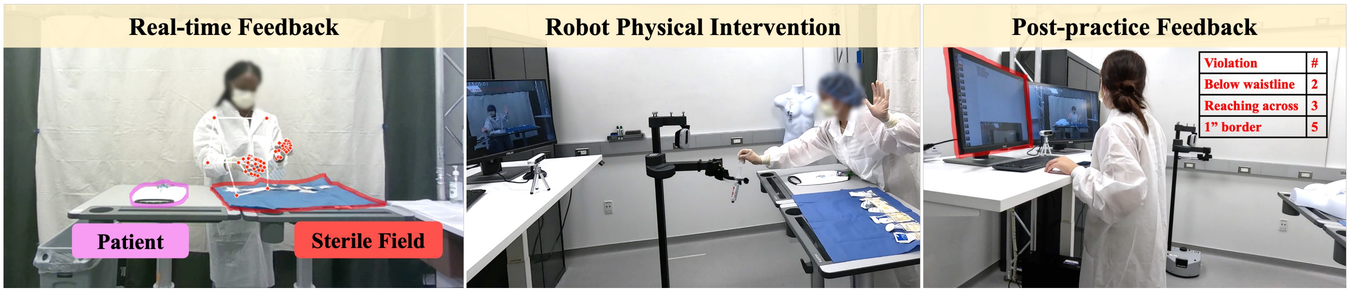

Fig. 2. Nurses practicing dressing change procedures with ASTRID in a training environment (also referred to as simulation lab in the clinical community). ASTRID offers (left) real-time guidance, (middle) physical interventions, and (right) post-practice feedback to help nurses master "principles of sterile technique" for preventing healthcare-associated infections.

nurse model [84, 25, 96]. ASTRID aims to complement this training by enabling nursing students, nursing residents, and early-career nurses (collectively referred to as nursing students in this paper) to practice and improve their skills, even without the presence of an expert nurse.

We, a cross-disciplinary team of roboticists and nursing professionals, designed ASTRID through a two-year participatory design process. First, through exploratory brainstorm sessions and requirement capture (Sec. III), we identified system requirements including necessary perception, manipulation, interaction, and tutoring capabilities. Guided by these requirements, we engaged in iterative prototype development (Sec. IV). As depicted in Fig. 1, ASTRID is realized using the Stretch mobile manipulator [69], an off-board depth camera [65], and a computer. Using its perception, ASTRID monitors students as they practice dressing changes on simulated patients (Fig. 2-left), providing real-time feedback when actions that could lead to infections are detected. Using its mobile manipulation, the robot simulates scenarios (Fig. 2-middle) that are associated with increased likelihood of human errors, such as interruptions. After each training session, ASTRID provides a summary report via the computer monitor (Fig. 2-right), enabling students to review their performance and identify areas for improvement. These features aim to complement the training provided by expert nurses, who may not always be available.

We evaluated ASTRID in a human-subject feasibility study with nine nurses (Sec. V). Results demonstrate that ASTRID can detect student errors almost as accurately as a nursing instructor, indicating its potential as an effective tutor. Additionally, subjective questionnaires reveal that participants find the system useful. Finally, the evaluations suggest several promising directions for both foundational research in robotics and their practical applications in nursing education.

II. Related Work

We review relevant literature on nursing, robotic tutors, and participatory design that informs our research. Figs. 3 and 4 summarizes the prior work in the area of robots in nursing and robotic tutoring systems, and illustrates the unique contribution of our work.

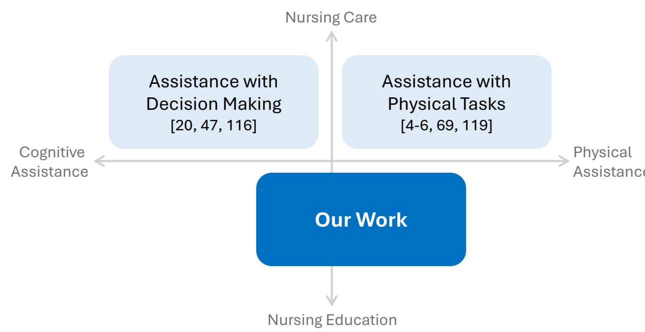

Fig. 3. Prior work in *robotics for nursing* focuses on providing nurses with assistance in decision making and physical tasks in nursing care. In contrast, this paper focuses on robotic assistance in nursing education.

A. Robotics for Nursing

In response to nursing shortages and growing workloads, "robotics for nursing" has become a vibrant research area [72, 20, 5, 67, 6]. Most of this work has focused on robots that assist nurses in homes and hospitals, with little work on robotics for nursing education.

1) Nursing Care: Robotic assistants, including commercially available products, are being developed to support nurses [50, 121, 124]. Pilot studies have shown success in using robots for tasks such as fetching supplies and disinfecting rooms [83, 4, 133]. As illustrated in Figure 3, existing systems provide a mix of cognitive and physical assistance, primarily focusing on nursing care. In contrast, our work centers on nursing education. While our focus differs, our approach is informed by robots designed for nursing care.

2) Nursing Education: Nursing research highlights the growing need for technology in nursing education [63, 107, 37, 38], with its role evolving rapidly since the COVID-19 pandemic. This shift has seen increased use of videos [43, 8, 35, 41] and simulations [32, 125, 136]. Closer to our focus, humanoid robot-patients [38] and telepresence robots are being explored to make nursing education more accessible [1]. However, to our knowledge, the potential of robotic tutors in nursing remains untapped and ASTRID is the first-of-its-kind robotic tutor for nursing.

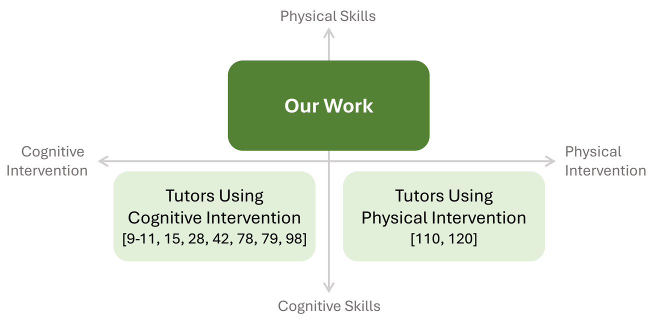

Fig. 4. Prior work on robotic tutors primarily focuses on cognitive skill training. In contrast, we explore the role of robotic tutors in physical skill training, integrating both cognitive and physical interventions to enhance physical skill execution.

B. Robotic Tutors

Intelligent tutoring systems have been developed for various learning environments, including K-12 education, corporate training, and medicine [103, 104, 94, 74, 59, 2, 3]. More recently, robotic tutors have emerged that offer additional benefits such as enhanced student interaction, improved learning outcomes, and increased trust and engagement [81, 15, 98, 10, 30, 9, 11, 44, 102, 82], by leveraging their physical presence and interaction capabilities.These systems allow students to practice and learn even when teachers are unavailable, aiming to enhance personalization and accessibility of instruction. Most robotic tutors, as illustrated in Figure 4, rely primarily on conversational instructions and lack mobile manipulation capabilities. Research in human-robot interaction (HRI) has looked using a physical intervention to improve human learners' cognitive skills such as problem solving [120] and knowledge in circuit design [110]. In contrast, our work uses a robot's perception and mobile manipulation capabilities to assess and enhance students' physical skill execution.

C. Participatory Design

Participatory design is a collaborative process that involves users in the process of designing a new service or technology. It combines the knowledge of the users with the skills of system designers, ensuring the inclusion of user needs in the design process [56, 21]. A participatory design process usually starts with reciprocal learning in which users and designers learn about each others' roles: designers learn about work practices from users and users learn about technical constraints from designers [28]. While many participatory design methods exist [116, 93], the most widely used methods include: participatory design workshops, collaborative focus groups, prototyping, visits to other institutions, and usability testing [70, 91]. Participatory design has found success in design of robots, including those designed for healthcare and tutoring [16, 78, 122, 132]. Informed by these works, we utilize a combination of participatory design methods to capture nurses' needs, develop prototypes collaboratively with nurses, and evaluate our prototype through a human-subject study.

III. Requirement Capture

To design the robotic tutor, we adopted a three-step participatory design process: requirement capture, prototype design, and feasibility study. This section focuses on the first step, aimed at first identifying which nursing skills would benefit most from robotic tutors and then determining the design requirements for the robotic tutor. We achieve these aims through an exploratory phase, which are followed by focused interviews with stakeholders.

A. Exploratory Phase

We believe robotic tutors have the potential to assist in a *variety* of nursing education settings. To identify the most suitable setting for the first such tutor, our cross-disciplinary team began with internal brainstorming sessions. These sessions were also crucial for aligning the team. Roboticists attended nurse training sessions to learn about clinical practices and build rapport with nursing professionals—an essential step for this interdisciplinary effort. Two early tutor prototypes were developed, which helped refine the research questions and assess feasibility. Through this exploration, the central line dressing change (CLDC) emerged as a key focus due to it being a frequently-performed procedure, need for periodic training due to occurence of preventable human errors, and potential for robotic assistance.

CLDC is a crucial step in maintaining the sterility and preventing infection of a central venous catheter or central line, a medical device used to deliver fluids, medications, or nutrition directly into a large vein. Maintenance of the central line is complex and life-critical, protecting patient against infections [51, 58]. One devastating complication during CLDC is the central line-associated bloodstream infection (CLABSI), which accounts for 17% of the almost one million healthcare-associated infections per year [48]. Fortunately, CLABSIs are preventable with meticulous nursing care and adherence to established protocols [84, 25, 96]. These safety protocols, referred to as the "principles of sterile technique," outline essential rules for maintaining sterility during dressing changes. In this paper, we focus on nursing skills corresponding to four key rules which are illustrated in Fig. 5.

B. Focused Interviews

Having identified a suitable nursing setting, we turned to defining the design requirements for the robotic tutor. We conducted focused interviews with three stakeholder groups: nursing students, experienced nurses, and nurse educators. Further details about the participant recruitment methodology are provided in the Appendix.

1) Methodology: The interview protocol was approved by Rice University IRB and structured in three parts, each addressing a specific goal. The first part focused on understanding participants' experiences with CLDC and the challenges they face in adhering to the sterile technique. The second part explored current training methods, asking if the challenges identified were addressed and what participants liked or disliked about existing methods. The third part

Fig. 5. Four prohibited behaviors during a sterile procedure, such as the central line dressing change. Once a sterile field is established (the area shown in blue), nurses need to maintain sterility by avoiding potential contamination. In particular, a nurse must keep hands above their waistline and the sterile field within vision at all times. Hands, with sterile gloves on, must not touch anything non-sterile such as the 1-inch border of the sterile field.

brainstormed potential solutions, starting with open-ended discussions about new training aids and then focusing on robotic tutors. Participants were shown a video of an early prototype of the tutoring system and asked for feedback on its usefulness and improvements. This prototype included human pose estimation (via MediaPipe) and provided visual alerts as on-screen text. The Appendix includes the list of interview questions and a link to the video of the prototype.

2) Participants: Ten participants were interviewed, including three nursing students, three bedside nurses, and four nurse educators. Participants' ages ranged from 20 to 49 years. The experienced nurses and nurse educators had between 6 and 16 years of experience as nurses, with an average of 11.3 years.

3) Results: The focused interviews led to three findings

- *All participants rated CLDC to be highly important; nursing students found maintaining the sterile field challenging.* When asked to rate how challenging CLDC was in their experience, four out the ten participants, including all nursing student participants, rated the procedure as challenging (> 4 on a scale of $1-7$); three rated it as neutral ($= 4$); and three rated it as not challenging (< 4). However, all participants pointed out that CLDC could become extremely complex when compounded with accompanying real-world factors such as disruptions and interruptions resulting from unexpected movements from the patients, other patients calling the nurse, or family members asking questions during the procedure.
- *Current training methods do not emphasize real-world factors.* Nursing students first learn about CLDC and sterile technique in nursing school. However, they typically do not receive hands-on practice until they come to hospitals for internships, usually in their final year of undergraduate study. During the classroom learning (nursing schools, orientations, and training sessions), instructors focus on the basics of CLDC (i.e., the step-by-step procedure), and do not emphasize the accompanying real-world factors that introduce complexity to CLDC.
- *New training aids, with specific features, can assist in acquisiting of nursing skills.* We asked the participants to brainstorm new training aids that could help nurses learn and practice the skills required for CLDC. Fig. 6 summarizes the features brainstormed by the participants, labeled with the number of groups (out of 3) that mentioned the feature. All groups supported a system that monitors nurses and provides feedback on medical errors. However, preferences for warnings and feedback varied. For real-time feedback, participants debated audio vs. visual alerts and ultimately agreed on both to accommodate different preferences. They emphasized clear, specific explanations of errors over generic beeping sounds to avoid alarm fatigue. For post-practice feedback, students preferred a log sheet with timestamps and nurse actions, while educators suggested a checklist tracking rule violations. Participants also recommended training aids that simulate real-world disruptions and interruptions. Students and bedside nurses further proposed different training levels based on experience. Lastly, participants had no specific preferences regarding the robotic tutor's appearance.

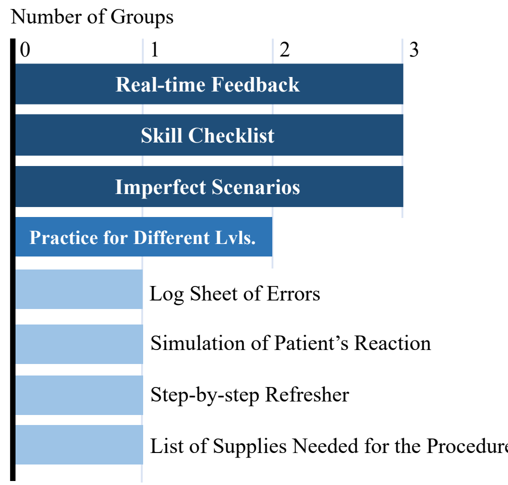

Fig. 6. Features suggested by the participants for new training aids.

C. System Requirements

Based on the findings of the exploratory phase and focused interviews, we distilled six key requirements (**Rx**) for a system that assists in CLABSI-prevention training. It needs to:

R1. detect compliance with the sterile technique;
R2. provide task-time guidance to facilitate skill acquisition;
R3. provide summary feedback for efficient training review;
R4. simulate scenarios that increase the risk of violations;
R5. be perceived as useful by nursing students; and
R6. be perceived as engaging by nursing students.

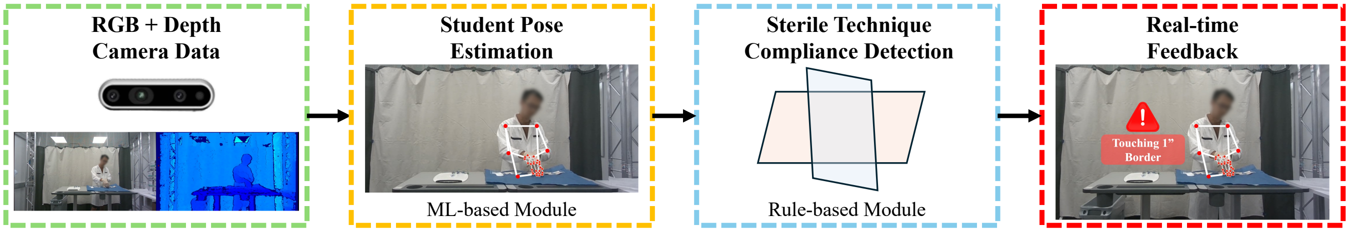

Fig. 7. Overview of ASTRID's system architecture for providing real-time guidance regarding the sterile technique to nursing students.

IV. PROTOTYPE DESIGN

Guided by the system requirements, we design ASTRID: the Automated Sterile Technique Review and Instruction Device shown in Fig. 2. It is intended for use in training environments. In this section, we detail its key features and their implementation.

A. System Overview

We begin by translating stakeholder requirements into the core capabilities the robotic tutor must possess:

- To detect nurses' compliance with sterile technique (**R1**): the robotic tutor requires cameras and computer vision algorithms to track nurses, perceive the environment, and assess compliance in real time. The perception and reasoning modules must operate with low latency to provide immediate feedback.
- To provide guidance during and after the task (**R2**, **R3**): the robot must generate nursing-specific feedback and communicate it effectively through audiovisual channels. The guidance must be both accurate and engaging to meet **R5** and **R6**, respectively.
- To simulate risk-inducing scenarios (**R4**): The system needs conversational and mobile manipulation capabilities to generate both verbal and physical interventions, such as interruptions and distractions.

These capabilities necessitate a robot with vision-based perception, mobile manipulation, and verbal interaction. Hence, to realize ASTRID, we build upon the Stretch mobile manipulator due to its human-safe design, onboard camera, and established use in healthcare robotics [69]. During requirement capture, a prototype demonstration using Stretch received positive feedback on nurses' comfort with the platform. To realize remaining capabilities, we augmented Stretch with additional hardware (an off-board depth camera [65] and a computer) and software modules for perception, reasoning, and interaction. Through the architecture in Fig. 7, ASTRID monitors nursing students as they practice central line dressing change, provides real-time feedback on sterile technique compliance, simulates error-prone scenarios, and generates a summary report for performance review.

B. Detecting Student Compliance with the Sterile Technique

Illustrated in Fig. 5, ASTRID considers four key principles of sterile technique once the sterile field is established. As nursing students practice the CLDC procedure, ASTRID monitors them and detects compliance with these principles as detailed next.

1) Sterile Field Detection: Each training session begins with ASTRID detecting the sterile field, which corresponds to the sterile drape, visible as the blue area on the table in Fig. 2. The drape is often uneven, irregularly shaped, and contains supplies inside pouches, making detection complex. To ensure robust detection, ASTRID utilizes a manual calibration method.[1] In particular, before training begins, the user is shown a real-time image of the training setup on a monitor and marks the four corners of the sterile drape using a mouse. This process takes less than 10 seconds to complete. Our geometry-based software, developed in-house using OpenCV2 [23] and MediaPipe [87], then uses the pixel positions and corresponding depth values to identify the edges of the sterile field and construct a 3D model of it. Since the sterile field remains static during the procedure, calibration is required only once. In our evaluations (Sec. V), the experiment proctor performs this calibration.

2) Student Pose Estimation: During a training session, ASTRID monitors the nursing student using its camera to estimate their pose. The pose estimation module is built using MediaPipe [87]. In particular, pose estimation is achieved using a series of pre-trained models, where the first stage detects human bodies in an RGB frame, and the second stage locates key landmarks on the hands and body. The hand estimation model identifies 21 landmarks, while the pose estimation model tracks 33, with the most relevant for our application being the shoulders, elbows, wrists, and hips. After the pose estimation locates the key landmarks, it returns the pixel coordinates of the landmarks. We then create a 3D environment reconstruction by projecting the landmarks back to the 3D space. This is to integrate the sterile field information with human pose estimation to enable compliance detection, described next. One unique challenge of our use case is that the nurse wears face masks, hairnets, and gloves, which obscure parts of the body, and often stands behind tables, limiting visibility and making pose estimation challenging. To tackle this, we set the visibility (a hyperparameter of MediaPipe) of these landmarks to 0 to enhance robustness of this automated pose estimation.

3) Sterile Technique Compliance Detection: Next, ASTRID uses the sensed landmarks of the student's pose and the sterile drape to determine compliance with the four principles of sterile technique. This compliance detection is achieved through a geometric rule-based module. The rule-based module partitions the 3D space into sterile and non-sterile regions, it calculates

[1] We also explored methods that do not require manual calibration. However, we found that they were reliable only when the drape was flat and free of items. Future work could explore improving automated detection of the sterile drape via interactive machine learning.

two key planes based on the sterile drape: the bottom plane, fitted along the drape, and the front plane, defined perpendicular to the bottom plane at the front-most edge (relative to the student). Finally, it applies the following geometry-based rules to check for four non-compliant behaviors:

- *hands below waistline*: if wrist-landmarks are below the sterile drape (i.e., the bottom plane).
- *reaching across the sterile field*: if any human landmark is ahead of the the sterile drape (i.e., the front plane).
- *turning back to the sterile field*: if the vector from the left to right shoulder faces away from the sterile drape.
- *touching the one-inch border*: if the fingertip-landmarks are within one-inch of the boundary of the sterile drape.

C. Providing Feedback to Students

Along with detecting sterile technique compliance, ASTRID offers feedback both during and after the training session.

1) Task-Time Feedback: Guided by the focused interviews, ASTRID uses both visual and audio channels to provide task-time feedback. As shown in Fig. 2-middle, during the training session, the nursing student can see their pose on the computer screen with real-time skeletal tracking. Fig. 2-left provides a snapshot of this screen. If ASTRID detects a non-compliant behavior, it alerts the nurse both visually with red text on the screen and aurally via pre-recorded audio. The visual and audio alerts have the same message, "Warning: <*specific non-compliant behavior*> (e.g., hands below waistline)."

2) Post-Practice Feedback: At the end of the training session, ASTRID provides the nursing student with tools to quickly review their practice using its screen (Fig. 2-right). First, each session is recorded with skeletal tracking, date, time, and warnings, allowing the nurse to see what they did right or wrong. Second, if the nurse prefers not to review the entire recording, ASTRID saves key frames when non-compliant behaviors are detected. Lastly, ASTRID generates a PDF report summarizing how many times each rule was broken, along with the screenshots of non-compliant behavior and associated warnings.

D. Simulating Challenging Nursing Scenarios

Guided by the findings of focused interviews, we design ASTRID to offer three levels of training – novice, intermediate, and advanced – which increase in complexity, simulating challenging real-world scenarios that nurses may encounter during dressing change procedures. A video demo of these scenarios is available at http://tiny.cc/rss-2025-astrid.

1) Novice: At this level, nurses practice the central line dressing change without distractions or interruptions, with all feedback features enabled. It is ideal for nursing students and nurses unfamiliar with the procedure.

2) Intermediate: This level additionally introduces distractions, simulating real-world scenarios where nurses may be interrupted by patients, family members, or other medical staff. Distractions are created by the robot, which moves around the environment, says pre-scripted greetings, and provides positive feedback "You are doing great! Keep going!" to the nurse.

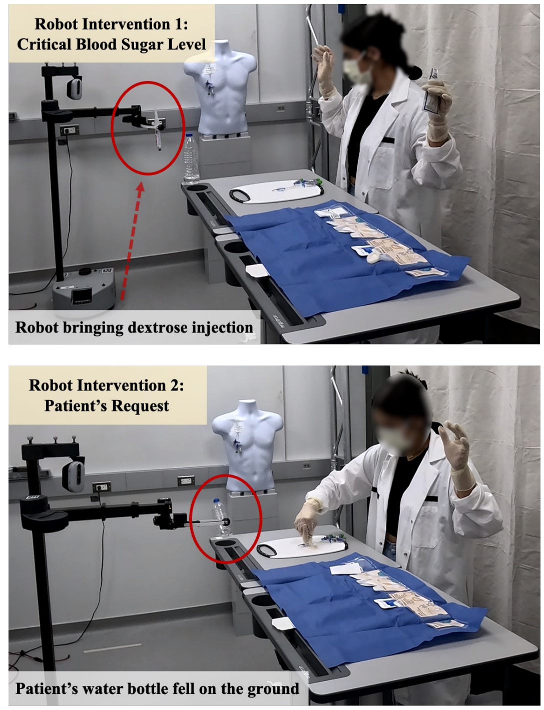

Fig. 8. ASTRID leverages mobile manipulation to create simulations of real-life scenarios.

The robot's paths are pre-designed giving considerations of nurse's safety and proxemics. During the feasibility study, the robot moves autonomously. The motion is programmed using *stretch_body*, a python API for Stretch.

3) Advanced: In the advanced level, ASTRID uses its mobile manipulation capability to simulate critical scenarios that require the nurse to apply their experience and judgment to determine the appropriate course of action. Currently, our prototype offers two such scenarios which were co-designed with nursing experts:

- *(Scenario #1)* ASTRID alerts the nurse, "Your patient's blood sugar is dropping below 54 mg/dL (milligrams per deciliter). I am bringing glucose." and brings a 50% dextrose (glucose) injection near the bedside tables (Fig. 8-top). Once the robot has reached the table, it alerts the nurse, "please take the glucose and give it to the patient." three times. If and when the nurse takes the dextrose injection, they reach across the sterile field, breaking the sterile field. This represents a life-critical scenario where the patient's blood sugar is dangerously low and could continue dropping, requiring the nurse to act quickly.
- *(Scenario #2)* ASTRID approaches the table and knocks over the patient's water bottle off the table, and alerts, "Oops, the patient water bottle fell on the floor, please pick up the water bottle and put it back." This scenario represents a non-emergency but common interruptions, e.g., where a patient drops something and asks the nurse to pick it up (Fig. 8-bottom). In such cases, experienced nurses would typically inform the patient that they will retrieve the item after the procedure.

The choice of these physical interventions was informed by challenges described by stakeholders during the focused interviews, direct observations on hospital floors, and discussions with nursing educators. These scenarios were refined to ensure they aligned with professional nursing practices, met our design goals, and were technically feasible to implement on a robot, resulting in the specific interventions described above. In both scenarios, the nurse must carefully reason through their actions. Following the robot's suggestion may lead to a violation of sterile technique, but in some cases (e.g., the first scenario) it may be necessary to prioritize patient health. Similar to the implementation in the Intermediate level, all robot's verbal and physical interactions are pre-programmed using the software package *stretch_body*, enabling the robot to behave autonomously during training sessions.

V. Feasibility Study

We conducted a feasibility study to evaluate ASTRID.[2] This IRB-approved study involved nine participants – seven recent graduates or nurses with two years or less of experience and two experienced nurses from the requirement capture interviews. The experiment was held in the training environment shown in Fig. 2, an artificial rendering of which is depicted in Fig. 1.

A. Materials

The experiment site was set up to mimic CLDC scenarios, with the same level of fidelity typically used in nursing education. The setup included a simulated patient, ASTRID, a table for performing the dressing change, medical supplies, and a GoPro camera. The depth camera and monitor were placed on a table in front of the nurse. The GoPro was used to record the experiments and was not part of ASTRID. One of the authors served as the experiment proctor.

B. Procedure

The experiment consisted of three parts: an introduction, dressing change procedures, and a post-experiment review.

1) Introduction: The session began with a greeting, followed by an explanation of the study's purpose, procedure, participant rights, and potential risks and benefits. Participants provided informed consent and completed a demographic survey on-site.

2) Central Line Dressing Change Procedures: Participants were asked to perform CLDC on the simulated patient four times (referred to as tests), each with a different setup and purpose.

- **Test 0** (Pre-test) involved participants performing a CLDC without warnings or interventions from ASTRID. This allowed them to familiarize themselves with the setup and enabled necessary calibration.
- **Test 1** implemented the intermediate level of ASTRID. Participants received real-time audio-visual warnings, in cases when ASTRID detected non-compliant behaviors. Additionally, to simulate real-life distractions, the robot moved around the room and said pre-scripted statements.
- **Test 2** implemented the advanced level of ASTRID. The participants continued to receive real-time warnings. Additionally, they had to respond to the two interruption scenarios described in Sec. IV-D3.
- **Test 3** implemented the novice level of ASTRID and involved the proctor instructing participants to deliberately perform non-compliant behaviors (e.g., dropping hands below the waistline). This test was included to evaluate ASTRID's detection capability, in case non-compliant behaviors were not observed in earlier tests.

After each test, the participants reviewed the summary report generated by ASTRID. The summary consists of the video recording of the test, snapshots of each detected non-compliant behavior, and a PDF report with a summary of how many times the participant broke each of the four rules along with the images of the mistakes.

3) Post-Experiment Review: After the participant completed all four tests, they were asked to complete a post-experiment survey administered on an on-site computer. Upon completing the survey, the experiment proctor conducted a brief interview to better understand the participant's experience and solicit suggestions for ASTRID's future iterations and usage.

C. Measures

The feasibility study assessed whether ASTRID met the design requirements (**R1**–**R6**) using a combination of objective and subjective measures. The post-experiment review also gathered participatory design feedback for future robotic tutors.

*1) Measures for **R1**:* To evaluate ASTRID's ability to detect student compliance with sterile technique, we compared its performance to that of a nurse educator. A nurse educator with 20+ years of experience reviewed and annotated the video recordings of the training sessions to establish the ground truth for non-compliant behaviors. Annotations were made using the Behavioral Observation Research Interactive Software (BORIS) [49], with the experiment proctor assisting in data entry. For analysis, we sampled the training sessions at 1Hz, treating each second as an instance for evaluation. Each instance was categorized as true positive (TP), false positive (FP), true negative (TN), or false negative (FN), based on ASTRID's detection compared to the expert annotations. For example, an instance was marked as FP if ASTRID detected a violation but the expert did not. Fig. 9 provides additional examples.

*2) Measures for **R2**–**R4**:* To evaluate **R2**–**R4**, the post-experiment survey included seven statements evaluating the usefulness of ASTRID's seven features (Fig. 10). Participants rated each feature on a 5-point discrete visual analog scale (DVAS), from not useful at all (1) to extremely useful (5).

*3) Measures for **R5**–**R6**:* The post-experiment survey included two established surveys, adapted for the experimental context, to assess the ASTRID's perceived usefulness [39] and user engagement [95]. The list of survey questions is provided in the appendix. The perceived usefulness survey included statements such as "Practicing with [technology] would enhance my overall job performance." Participants rated these statements on a 5-point discrete visual analog scale,

[2] We refer the reader to the supplementary material for video snippets and resources to support the reproducibility of this study.

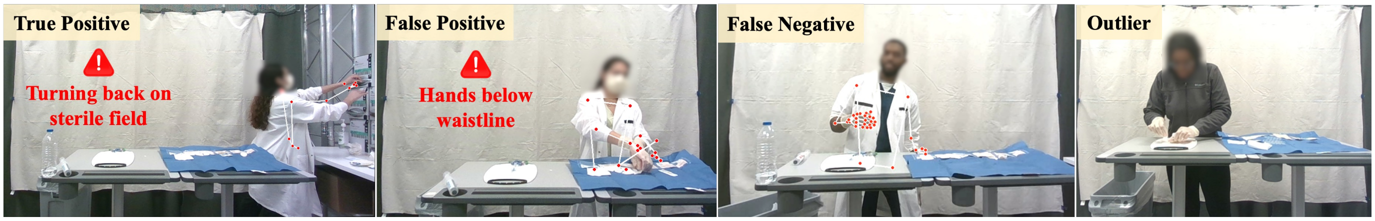

Fig. 9. Examples of ASTRID's compliance detection from the feasibility study: (True Positive) ASTRID correctly detects the nurse turning their back to the sterile field; (False Positive) ASTRID mistakenly identifies the hands as below the waistline due to occlusion; (False Negative) ASTRID misses the left hand below the waistline as it is hidden behind the back; and (Outlier) the table's higher height makes it challenging to detect the sterile field and compliance.

TABLE I
PERFORMANCE: STERILE TECHNIQUE COMPLIANCE DETECTION

Metrics	Results	Calculation
Accuracy	98.6%	(TP+TN) / (TP+TN+FP+FN)
Precision	95.5%	TP / (TP+FP)
Recall	83.5%	TP / (TP+FN)
$F1$ Score	0.89	2×Recall×Precision / (Recall+Precision)

ranging from extremely unlikely (1) to extremely likely (5). The user engagement survey included statements such as "My experience was rewarding." This scale also utilized a 5-point discrete visual analog scale: strongly disagree (1) to strongly agree (5).

D. Findings

Nine nurses participated in the feasibility study. Data from one participant ($P2$) was excluded as an outlier due to issues with the table height. Although we had opted for a height-adjustable table to accommodate participants of different height, the highest-level of the table was still below the participant's waistline. Of the remaining eight participants, each performed four tests, yielding 32 total tests. Two tests were not recorded properly due to data storage limitations. Additionally, two tests (pre-test and Test #1 of $P6$) were excluded because we noticed when the table was at its highest level, it was at the same level of the camera, making perception difficult. Though this was resolved when they performed Test #2 and Test #3. In total, we collected data from 28 effective tests, amounting to $5,719$ seconds (95.3 minutes) of video recordings.

Finding 1: ASTRID *demonstrates the potential to accurately detect student's compliance with the sterile technique.* Among the 5719 instances of data, 343 are found to be TP, 16 FP, 5376 TN, and 68 FN. This corresponds to an accuracy of 98.6% and an $F1$ score of 0.89. An $F1$ score above 0.9 is considered excellent, and an $F1$ score between 0.8 and 0.9 is considered good. Table I summarizes key metrics, demonstrating ASTRID's high accuracy in detecting both positives and negatives. While true positives are crucial for robotic tutoring, we wish to highlight that true negatives are equally important. They validate ASTRID's perception modules and (as indicated in the open-ended feedback) nursing students felt encouraged when they knew they did not make any mistakes. Together these results indicate that ASTRID satisfies requirement **R1**.

Finding 2: Participants perceive ASTRID *as highly useful for nursing education.* Through the perceived usefulness survey, detailed in the appendix, participants reported a mean score of $M = 4.70$ (out of a maximum of 5) and $SD = 0.41$ on the system's perceived usefulness. This result highlights the potential of ASTRID as a helpful tutoring system to improve nurses' skills and job performance in the future. Fig. 10 provides a granular view of ASTRID's perceived usefulness, highlighting participants' feedback on its individual features. While some features were rated more useful than others, both real-time and post-practice feedback were consistently rated as extremely useful.

Finding 3: Participants perceive ASTRID *as engaging and supportive.* Through the user engagement survey, also detailed in the appendix, participants reported a mean score $M = 4.84$ (out of 5) and $SD = 0.30$, indicating that the participants found ASTRID engaging and supportive. In the post-experiment interview, we asked participants whether practicing with ASTRID would improve nursing students' confidence in complying with the sterile technique. All participants unanimously answered "yes." One participant, who recently graduated from nursing school said

> I always feel very anxious because I am new to the job. But practicing with the robot has already made me feel better about my skills because now I know I did not make any mistakes... I also like the real-life scenarios. We did not see those in nursing school.

In addition to helping nursing students improve and gain confidence in their skills, participants noted that ASTRID could offer other benefits, such as being more readily available than experienced nurses, providing objective assessments and feedback, and allowing nurses to practice their skills without fear of judgment. These results and comments are especially encouraging, as they align with the requirements identified through initial interviews and validate our participatory design efforts to create a nurse-friendly robotic tutor.

E. Limitations

While our work involved a rigorous co-design process, iterative development, and human-subject evaluations, it also has limitations. Here, we acknowledge these to contextualize our contributions, highlight the challenges of systems research, and guide future work. First, while we engaged stakeholders from the outset, our evaluations were limited to nine nurses. This sample size aligns with user-centered design research,

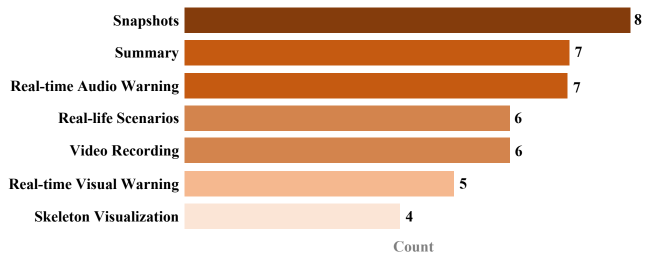

Fig. 10. Perceived usefulness of ASTRID's individual features: the number of participants (out of 8) that rated a feature to be *extremely useful*. While all features are perceived as useful, snapshots of contamination occurrences receive the highest ranking and skeleton visualization the lowest.

which suggests that 8–10 participants can uncover up to 80% of usability issues [79], including in healthcare technologies [75]. However, the small cohort can limit generalizability and our results should be considered proof-of-concept. A key challenge was recruiting novice nurses, whose demanding schedules made research participation difficult. Second, our user evaluations relied on established scales for perceived usefulness and user engagement, but we adapted them to the specific context of nursing education. These modifications were not separately validated. Lastly, while the robot operated autonomously, its functionality was constrained to the controlled training environment. For instance, the system was tested against a specific background. In real-world settings, the robot will need to adapt to a wider range of environments and users, requiring enhanced autonomy. As discussed in the next section, this underscores the need for foundational robotics research driven by real-world applications in nursing.

VI. Conclusion

We conclude by discussing the implications of our findings for both nursing education and foundational robotics research.

A. Implications for Nursing Education

1) Key Contributions: Our work introduces a novel technological aid for nursing education: robotic tutors. Through participatory design, we developed ASTRID, a robotic tutor for CLABSI-prevention training. ASTRID monitors student compliance with the sterile technique, providing real-time feedback and tools for performance review and improvement. It also offers multiple training levels and can simulate challenging scenarios which may be overlooked in nursing schools. In a feasibility study with 9 nurses, ASTRID reliably detected compliance with four key sterile technique principles and was perceived as useful and engaging. These results suggest that ASTRID can help provide new nurses with opportunities to practice their skills and receive immediate feedback, especially as current nursing shortages challenge the sustainability of the traditional nurse-to-nurse training model [47, 101, 63]. Further, our approach reaffirms the importance of involving nurses early on during the design and development of new technology.

2) Directions for Future Work: While ASTRID shows promise, we emphasize that it is a proof-of-concept system. We list suggest future directions for nursing research:

- ASTRID addresses four principles of the sterile technique; however, this technique is more comprehensive [84, 25, 96]. The system also focuses on a specific part of the dressing change procedure – after the nurse has already opened the dressing change kit and set up the sterile field. Early steps like opening the sterile packet and putting on gloves are not covered and should be addressed in future work. Participants also suggested expanding the tutor to other tasks like Foley catheter insertion [18], and high-sterility environments like operating rooms (OR) [118, 112, 106, 128].
- While ASTRID reliably detects sterile technique violations, it is not immune to errors. A risk is students becoming overconfident due to false negatives. While technological improvements can enhance detection accuracy, we believe integrating input from nursing instructors is crucial to address this challenge. Participants also emphasized the importance of human instruction, especially for those who began their education during COVID-19. Thus, ASTRID should complement broader nursing education frameworks rather than function as a standalone tool. It is important to highlight that ASTRID is designed to augment traditional nursing training and intended for use alongside instructors. Future work should evaluate ASTRID through formal A/B comparison against a control condition where nurses practiced without robotic support. Additionally, our current work focused on student-robot interactions, with instructor input limited to its design and evaluation. Future work should explore the instructor-student-robot triad to understand how robots can best support existing teaching methods.

B. Implications for Robotics

1) Key Contributions: Within robotics, our work makes three key contributions:

- *Introducing nursing education as a new robotics domain*: We identify nursing education as a field in need of transformative solutions and demonstrate the potential of robotics in this space through the design, development, and evaluation of ASTRID. Our work serves as an initial testbed for evaluating robotics technologies in nursing education and reveals new research directions in task and motion planning, perception-aware motion planning, conversational robots, and robotic tutors.
- *Advancing robotic tutoring with a focus on physical skills*: Unlike most robotic tutors, which focus on cognitive learning through conversational interactions (e.g., math tutoring), ASTRID is designed for physical skill acquisition. This shift necessitates perception-driven performance evaluation and physical interventions via mobile manipulation. We see this as a small but important step toward expanding robotic tutoring beyond screen-based or purely conversational approaches.

- *Reaffirming the value of participatory design in robotics*: Our work reinforces the importance of participatory design in developing robotics systems. By actively involving stakeholders throughout the process, we ensure that the technology is aligned with real-world needs, further supporting user-centered approaches in robotics research.

2) Directions for Future Work: Our systems-driven investigation also reveals directions for foundational research in robotics, which have implications beyond nursing:

- *Rule-Based Detection and Alternatives:* The rule-based perception pipeline of ASTRID offers simplicity and reliability (advantages that are valuable in early-stage research) but limits generalizability. Future work should focus on enhancing detection using more robust pose estimation, object detection, and action recognition techniques. In ongoing work, we are evaluating alternatives to MediaPipe for improved pose estimation, and plan to integrate object detection models (e.g., YOLO [105], SSD [85]) and action recognition models (e.g., I3D [29] and SlowFast [46]) to enhance generalizability.
- *Perception-aware Robot Planning*: During evaluations, ASTRID struggled with taller participants, partly due to the use of a static off-board camera. Raising the camera reduces visibility of the lower body, while lowering it obstructs key areas like the far side of the sterile field. Future work should consider development of more robust perception systems for detecting nursing activities, by leveraging advances in vision [34, 52, 119, 60, 127, 89] and mobile perception [45, 64, 137, 126]. Although there is a vast literature on perception-aware motion generation and active perception, we have identified additional constraints that prevent the seamless application of existing methods, such as the generation of robot motion for object manipulation that maximizes the nurse monitoring, especially for high degree-of-freedom robots. These constraints also extend to task planning and motivate the need for novel methods for *perception-aware task and motion planning*.
- *Multimodal Human-(Robotic Tutor) Interaction*: Several nurses attempted to converse with ASTRID when it greeted them or issued verbal warnings, but ASTRID relies on pre-scripted language and lacks the ability for free-form, turn-taking conversations. Adding such features could enhance the tutoring. However, if generative or large language models are used, their limitations must be carefully considered [17, 97, 138, 22, 71, 129]. Combining multiple intervention types also offers exciting potential; our work takes a step in this direction by incorporating physical interventions, though limited to pre-defined tasks. Tighter integration of perception, mobile manipulation, and conversation could significantly enhance future robotic tutors across domains.
- *Calibrating Trust in Robotic Tutors*: As robotic tutors become more capable, trust calibration is a critical concern to prevent students from over-relying on these systems [40, 134, 135, 80, 73, 53]. This is especially important for robotic tutors, as their teaching role may lead students to inherent trust them more than robotic assistants or peers [19, 24, 114, 36, 86, 111]. One way to address this is by explaining the robot's capabilities and limitations to students and involving instructors in the process [57, 12, 76, 13, 14, 31, 77, 55, 61, 117, 123, 131, 99, 27, 108, 109, 92, 100]. Indeed, our participatory design findings suggest that allowing educators to customize robotic tutors through end-user programming will be essential for their real-world effectiveness [33, 7, 139, 88, 68].

We conclude by emphasizing the need for cross-disciplinary collaboration in use-inspired robotics systems research. Our research reaffirms that developing robotics systems requires inputs from domain experts, use of participatory design methods, and expertise from robot developers. This integrated approach is key to developing safe and responsible human-centered robotics.

VII. Ethical Impact Statement

This work presents ASTRID, a robotic tutor designed to support nursing education and reduce preventable infections through improved training. Developed through a participatory design process with nursing professionals, ASTRID aims to **complement** human instruction. While it offers real-time and post-practice feedback, we caution against overreliance due to potential perception errors. All user studies were conducted under IRB approval, with informed consent and attention to data privacy. Broader deployment should ensure equitable access, human instructor's oversight, and mechanisms to calibrate user trust in the robotic tutor.

VIII. Summary of Supplementary Material

Please see the Appendix for further details on:

- Methodology for recruiting participants;
- Questions used during the focused interviews;
- Statements used during the feasibility study to measure ASTRID's perceived usefulness; and
- Statements used during the feasibility study to measure user engagement during the experiment.

Additionally, a video demonstration of ASTRID is available at http://tiny.cc/rss-2025-astrid.

Acknowledgments

This research was supported by the National Science Foundation through Award 2326390 and Rice University funds. We acknowledge the inputs of nursing students, nurses, and nursing instructors, who were critical to this participatory design effort. Lastly, we thank the anonymous reviewers for their thoughtful feedback and constructive suggestions.

References

[1] Alham Abuatiq, Robin Brown, Christina Plemmons, Beth Walstrom, Cassy Hultman, Danielle Currier, Marie Schmit, Valborg Kvigne, Leann Horsley, and Heidi Mennenga. Nursing faculty and students' satisfaction with telepresence robots during the covid-19 pandemic. *Nurse Educator*, 47(2):E39–E42, 2022.

[2] Moh'd Abuazizeh, Thomas Kirste, and Kristina Yordanova. Computational state space model for intelligent tutoring of students in nursing subjects. In *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, pages 1–7, 2020.

[3] Moh'd Abuazizeh, Kristina Yordanova, and Thomas Kirste. Affect-aware conversational agent for intelligent tutoring of students in nursing subjects. In Alexandra I. Cristea and Christos Troussas, editors, *Intelligent Tutoring Systems*, pages 497–502, Cham, 2021. Springer International Publishing. ISBN 978-3-030-80421-3.

[4] Shamsudeen Abubakar, Sumit K Das, Chris Robinson, Mohammed N Saadatzi, M Cynthia Logsdon, Heather Mitchell, Diane Chlebowy, and Dan O Popa. Arna, a service robot for nursing assistance: System overview and user acceptability. In *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, pages 1408–1414. IEEE, 2020.

[5] E Ackerman. Diligent robotics bringing autonomous mobile manipulation to hospitals. *IEEE Spectrum*, 2018.

[6] E Ackerman. Akara robotics turns turtlebot into autonomous uv disinfecting robot. *IEEE Spectrum*, 2020.

[7] Gopika Ajaykumar, Maureen Steele, and Chien-Ming Huang. A survey on end-user robot programming. *ACM Computing Surveys (CSUR)*, 54(8):1–36, 2021.

[8] Naseem Saeed Ali and Bindu John. Examining the efficacy of online self-paced interactive video-recordings in nursing skill competency learning: seeking preliminary evidence through an action research. *Medical Science Educator*, 29:463–473, 2019.

[9] Patrícia Alves-Oliveira, Srinivasan Janarthanam, Ana Candeias, Amol Deshmukh, Tiago Ribeiro, Helen Hastie, Ana Paiva, and Ruth Aylett. Towards dialogue dimensions for a robotic tutor in collaborative learning scenarios. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 862–867. IEEE, 2014.

[10] Patrícia Alves-Oliveira, Tiago Ribeiro, Sofia Petisca, Eugenio Di Tullio, Francisco S Melo, and Ana Paiva. An empathic robotic tutor for school classrooms: Considering expectation and satisfaction of children as end-users. In *Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings 7*, pages 21–30. Springer, 2015.

[11] Patrícia Alves-Oliveira, Pedro Sequeira, and Ana Paiva. The role that an educational robot plays. In *2016 25th IEEE International symposium on robot and human interactive communication (RO-MAN)*, pages 817–822. IEEE, 2016.

[12] Dan Amir and Ofra Amir. Highlights: Summarizing agent behavior to people. In *Proceedings of the 17th international conference on autonomous agents and multiagent systems*, pages 1168–1176, 2018.

[13] Ofra Amir, Finale Doshi-Velez, and David Sarne. Summarizing agent strategies. *Autonomous Agents and Multi-Agent Systems*, 33:628–644, 2019.

[14] Sule Anjomshoae, Amro Najjar, Davide Calvaresi, and Kary Främling. Explainable agents and robots: Results from a systematic literature review. In *18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019*, pages 1078–1088. International Foundation for Autonomous Agents and Multiagent Systems, 2019.

[15] Wilma A Bainbridge, Justin W Hart, Elizabeth S Kim, and Brian Scassellati. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics*, 3:41–52, 2011.

[16] Tony Belpaeme, Paul Vogt, Rianne Van den Berghe, Kirsten Bergmann, Tilbe Göksun, Mirjam De Haas, Junko Kanero, James Kennedy, Aylin C Küntay, Ora Oudgenoeg-Paz, et al. Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics*, 10:325–341, 2018.

[17] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623, 2021.

[18] Alexa Bianchi, Stephen W Leslie, and Gregory T Chesnut. Difficult foley catheterization. *StatPearls [Internet]*, 2023.

[19] Chris Birmingham, Zijian Hu, Kartik Mahajan, Eli Reber, and Maja J Matarić. Can i trust you? a user study of robot mediation of a support group. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8019–8026. IEEE, 2020.

[20] Richard Bloss. Mobile hospital robots cure numerous logistic needs. *Industrial Robot: An International Journal*, 2011.

[21] Susanne Bødker, Christian Dindler, Ole S. Iversen, and Rachel C. Smith. *What Is Participatory Design?*, pages 5–13. Springer International Publishing, 2022.

[22] Anya Bouzida, Alyssa Kubota, Dagoberto Cruz-Sandoval, Elizabeth W Twamley, and Laurel D Riek. Carmen: A cognitively assistive robot for personalized neurorehabilitation at home. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 55–64, 2024.

[23] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[24] Kimberly A Brink and Henry M Wellman. Robot teachers for children? young children trust robots depending on their perceived accuracy and agency. *Developmental Psychology*, 56(7):1268, 2020.

[25] Niccolò Buetti, Jonas Marschall, Marci Drees, Mohamad G. Fakih, Lynn Hadaway, Lisa L. Maragakis, Elizabeth Monsees, Shannon Novosad, Naomi P O'Grady, Mark E. Rupp, Joshua Wolf, Deborah Yokoe, and Leonard A. Mermel. Strategies to prevent central line-associated bloodstream infections in acute-care

hospitals: 2022 update. *Infection control and hospital epidemiology*, 2022.

[26] Tyler Bysshe, Yue Gao, Krysta Heaney-Huls, Jason Hockenberry, Lauren Hovey, Alison M. Laffan, Suhna Lee, David J. Murphy, and Elizabeth Watts. Estimating the additional hospital inpatient cost and mortality associated with selected hospital-acquired conditions. Technical report, Agency for Healthcare Research and Quality, 2017.

[27] Davide Calvaresi, Amro Najjar, Andrea Omicini, Reyhan Aydogan, Rachele Carli, Giovanni Ciatto, Yazan Mualla, and Kary Främling. Explainable and transparent ai and multi-agent systems. *LECTURE NOTES IN COMPUTER SCIENCE*, 14127:1–281, 2023.

[28] Erran Carmel, Randall D. Whitaker, and Joey F. George. Pd and joint application design: a transatlantic comparison. *Commun. ACM*, 36:40–48, 1993. URL https://api.semanticscholar.org/CorpusID:6947185.

[29] João Carreira and Andrew Zisserman. Quo vadis, action recognition? a new model and the kinetics dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4724–4733, 2017. doi: 10.1109/CVPR.2017.502.

[30] Ginevra Castellano, Ana Paiva, Arvid Kappas, Ruth Aylett, Helen Hastie, Wolmet Barendregt, Fernando Nabais, and Susan Bull. Towards empathic virtual and robotic tutors. In *Artificial Intelligence in Education: 16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013. Proceedings 16*, pages 733–736. Springer, 2013.

[31] Tathagata Chakraborti, Anagha Kulkarni, Sarath Sreedharan, David E Smith, and Subbarao Kambhampati. Explicability? legibility? predictability? transparency? privacy? security? the emerging landscape of interpretable agent behavior. In *Proceedings of the international conference on automated planning and scheduling*, volume 29, pages 86–96, 2019.

[32] Jie Chen, Jian Yang, Fen Hu, Si-Hong Yu, Bing-Xiang Yang, Qian Liu, and Xiao-Ping Zhu. Standardised simulation-based emergency and intensive care nursing curriculum to improve nursing students' performance during simulated resuscitation: a quasi-experimental study. *Intensive and Critical Care Nursing*, 46:51–56, 2018.

[33] Sonia Chernova and Andrea L Thomaz. Robot learning from human teachers. *Synthesis lectures on artificial intelligence and machine learning*, 8(3):1–121, 2014.

[34] Wongun Choi, Caroline Pantofaru, and Silvio Savarese. Detecting and tracking people using an rgb-d camera via multiple detector fusion. In *2011 IEEE international conference on computer vision workshops (ICCV workshops)*, pages 1076–1083. IEEE, 2011.

[35] Yeu-Hui Chuang, Fu-Chih Lai, Chia-Chi Chang, and Hsu-Tien Wan. Effects of a skill demonstration video delivered by smartphone on facilitating nursing students' skill competencies and self-confidence: A randomized controlled trial study. *Nurse education today*, 66:63–68, 2018.

[36] Cristina Conati, Oswald Barral, Vanessa Putnam, and Lea Rieger. Toward personalized xai: A case study in intelligent tutoring systems. *Artificial intelligence*, 298: 103503, 2021.

[37] Angelo Dante, Alessia Marcotullio, Vittorio Masotta, Valeria Caponnetto, Carmen La Cerra, Luca Bertocchi, Cristina Petrucci, and Celeste M Alfes. From high-fidelity patient simulators to robotics and artificial intelligence: A discussion paper on new challenges to enhance learning in nursing education. In *Methodologies and Intelligent Systems for Technology Enhanced Learning, 10th International Conference. Workshops: Volume 2*, pages 111–118. Springer, 2021.

[38] Angelo Dante, Carmen La Cerra, Vittorio Masotta, Valeria Caponnetto, Luca Bertocchi, Alessia Marcotullio, Fabio Ferraiuolo, Celeste M Alfes, and Cristina Petrucci. The use of robotics to enhance learning in nursing education: a scoping review. In *Methodologies and Intelligent Systems for Technology Enhanced Learning, 11th International Conference 11*, pages 217–226. Springer, 2022.

[39] Fred Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13:319–, 09 1989. doi: 10.2307/249008.

[40] Munjal Desai, Poornima Kaniarasu, Mikhail Medvedev, Aaron Steinfeld, and Holly Yanco. Impact of robot failures and feedback on real-time trust. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 251–258. IEEE, 2013.

[41] Barkha Devi, Bidita Khandelwal, and Mridula Das. Comparison of the effectiveness of video-assisted teaching program and traditional demonstration on nursing students learning skills of performing obstetrical palpation. *Iranian journal of nursing and midwifery research*, 24(2):118, 2019.

[42] Jonathan Dhaussy, Lucie Kemken, Marie-Thérèse Pugliese, Aline Forestier, and Sylvain Boloré. Using simulation to adapt nursing education to times of crisis: A scoping review during covid-19 pandemic. *Teaching and Learning in Nursing*, 2024.

[43] Georgia Ann Dinndorf-Hogenson, Carrie Hoover, Jodi Lisbeth Berndt, Bethany Tollefson, Jennifer Peterson, and Nichole Laudenbach. Applying the flipped classroom model to psychomotor skill acquisition in nursing. *Nursing education perspectives*, 40(2):99–101, 2019.

[44] Melissa Donnermann, Philipp Schaper, and Birgit Lugrin. Social robots in applied settings: A long-term study on adaptive robotic tutors in higher education. *Frontiers in Robotics and AI*, 9:831633, 2022.

[45] Davide Falanga, Philipp Foehn, Peng Lu, and Davide Scaramuzza. Pampc: Perception-aware model predictive control for quadrotors. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*,

pages 1–8. IEEE, 2018.
[46] Haoqi Fan, Yanghao Li, Bo Xiong, Wan-Yen Lo, and Christoph Feichtenhofer. Pyslowfast. https://github.com/facebookresearch/slowfast, 2020.
[47] Julie Fitzwater, Jeanette McNeill, Diane Monsivais, and Franchesca Nunez. Using simulation to facilitate transition to the nurse educator role: An integrative review. *Nurse educator*, 2021.
[48] Joseph D Forrester, Paul M Maggio, and Lakshika Tennakoon. Cost of health care-associated infections in the united states. *Journal of patient safety*, 2022.
[49] Olivier Pierre Friard, Marco Gamba, et al. Behavioral observation research interactive software (boris). 2016.
[50] Matthew Gombolay, Xi Jessie Yang, Bradley Hayes, Nicole Seo, Zixi Liu, Samir Wadhwania, Tania Yu, Neel Shah, Toni Golen, and Julie Shah. Robotic assistance in the coordination of patient care. *The International Journal of Robotics Research*, 37(10):1300–1316, 2018.
[51] Lisa A Gorski, Lynn Hadaway, Mary E Hagle, Daphne Broadhurst, Simon Clare, Tricia Kleidon, Britt M Meyer, Barb Nickel, Stephen Rowley, Elizabeth Sharpe, and Mary Alexander. Infusion therapy standards of practice, 8th edition. *Journal of infusion nursing: the official publication of the Infusion Nurses Society*, 2021.
[52] Chunhui Gu, Chen Sun, David A Ross, Carl Vondrick, Caroline Pantofaru, Yeqing Li, Sudheendra Vijayanarasimhan, George Toderici, Susanna Ricco, Rahul Sukthankar, et al. Ava: A video dataset of spatio-temporally localized atomic visual actions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6047–6056, 2018.
[53] Yaohui Guo and X Jessie Yang. Modeling and predicting trust dynamics in human–robot teaming: A bayesian inference approach. *International Journal of Social Robotics*, 13(8):1899–1909, 2021.
[54] Lisa M Haddad, Pavan Annamaraju, and Tammy J Toney-Butler. Nursing shortage. *StatPearls [Internet]*, 2020.
[55] Zhao Han, Daniel Giger, Jordan Allspaw, Michael S Lee, Henny Admoni, and Holly A Yanco. Building the foundation of robot explanation generation using behavior trees. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(3):1–31, 2021.
[56] Rex Hartson and Partha S. Pyla. *The UX Book*. Elsevier, 2012.
[57] Bradley Hayes and Julie A Shah. Improving robot controller transparency through autonomous policy explanation. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 303–312, 2017.
[58] Matthew A Hicks, Patrycja Popowicz, and Peter P Lopez. Central line management. In *StatPearls [Internet]*. StatPearls Publishing, 2023.
[59] Marjan Hospers, Erna Kroezen, Anton Nijholt, Rieks op den Akker, and Dirk Heylen. An agent-based intelligent tutoring system for nurse education. *Applications of Software Agent Technology in the Health Care Domain*, pages 143–159, 2003.
[60] Ming Hu, Lin Wang, Siyuan Yan, Don Ma, Qingli Ren, Peng Xia, Wei Feng, Peibo Duan, Lie Ju, and Zongyuan Ge. Nurvid: A large expert-level video database for nursing procedure activity understanding. *Advances in Neural Information Processing Systems*, 36, 2024.
[61] Sandy H. Huang, David Held, Pieter Abbeel, and Anca D. Dragan. Enabling robots to communicate their objectives. *Autonomous Robots*, 43(2), February 2019.
[62] Ronda G Hughes(ed.). *Patient Safety and Quality: An Evidence-Based Handbook for Nurses*. Agency for Healthcare Research and Quality (AHRQ), 2008.
[63] Kathleen M Huun and James E Slaven. Robotic telepresence and face-to-face collaborative nursing simulation: A correlational, cross-sectional study. *Clinical Simulation in Nursing*, 90:101525, 2024.
[64] Brian Ichter, Benoit Landry, Edward Schmerling, and Marco Pavone. Perception-aware motion planning via multiobjective search on gpus. In *Robotics Research: The 18th International Symposium ISRR*, pages 895–912. Springer, 2020.
[65] Intel. Intel realsense technology, 2024. URL https://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html.
[66] Minako Ito, Haruhiko Mitsunaga, and Toshiko Ibe. Survey of onboarding programs of hospitals for newly hired experienced nurses. *Journal of St. Luke's Society for Nursing Research (SLNR)*, 24, 2021.
[67] Mari Kangasniemi, Suyen Karki, Noriyo Colley, and Ari Voutilainen. The use of robots and other automated devices in nurses' work: An integrative review. *International journal of nursing practice*, 25(4):e12739, 2019.
[68] Ulas Berk Karli, Juo-Tung Chen, Victor Nikhil Antony, and Chien-Ming Huang. Alchemist: Llm-aided end-user development of robot applications. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 361–370, 2024.
[69] Charles C Kemp, Aaron Edsinger, Henry M Clever, and Blaine Matulevich. The design of stretch: A compact, lightweight mobile manipulator for indoor human environments. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 3150–3157. IEEE, 2022.
[70] Finn Kensing, Jesper Simonsen, and Keld Bodker. Must: A method for participatory design. *Human–Computer Interaction*, 13, 1998.
[71] Callie Y Kim, Christine P Lee, and Bilge Mutlu. Understanding large-language model (llm)-powered human-robot interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 371–380, 2024.
[72] Thomas E Kirschling, Steve S Rough, and Brad C Ludwig. Determining the feasibility of robotic courier

medication delivery in a hospital setting. *American Journal of Health-System Pharmacy*, 66(19):1754–1762, 2009.

[73] Bing Cai Kok and Harold Soh. Trust in robots: Challenges and opportunities. *Current Robotics Reports*, 1(4):297–309, 2020.

[74] C Koutsojannis, J Prentzas, and I Hatzilygeroudis. A web-based intelligent tutoring system teaching nursing students fundamental aspects of biomedical technology. In *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 4, pages 4024–4027. IEEE, 2001.

[75] Andre W Kushniruk and Vimla L Patel. Cognitive and usability engineering methods for the evaluation of clinical information systems. *Journal of biomedical informatics*, 37(1):56–76, 2004.

[76] Minae Kwon, Sandy H Huang, and Anca D Dragan. Expressing robot incapability. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 87–95, 2018.

[77] Isaac Lage, Daphna Lifschitz, Finale Doshi-Velez, and Ofra Amir. Exploring computational user models for agent policy summarization. In *IJCAI: proceedings of the conference*, volume 28, page 1401. NIH Public Access, 2019.

[78] Hee Rin Lee, Selma Šabanović, Wan-Ling Chang, Shinichi Nagata, Jennifer Piatt, Casey Bennett, and David Hakken. Steps toward participatory design of social robots: Mutual learning with older adults with depression. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17, page 244–253. Association for Computing Machinery, 2017. ISBN 9781450343367.

[79] James R Lewis. Sample sizes for usability studies: Additional considerations. *Human factors*, 36(2):368–378, 1994.

[80] Michael Lewis, Katia Sycara, and Phillip Walker. The role of trust in human-robot interaction. *Foundations of trusted autonomy*, pages 135–159, 2018.

[81] Daniel Leyzberg, Samuel Spaulding, Mariya Toneva, and Brian Scassellati. The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the annual meeting of the cognitive science society*, volume 34, 2012.

[82] Daniel Leyzberg, Aditi Ramachandran, and Brian Scassellati. The effect of personalization in longer-term robot tutoring. *ACM Transactions on Human-Robot Interaction (THRI)*, 7(3):1–19, 2018.

[83] Zhi Li, Peter Moran, Qingyuan Dong, Ryan J Shaw, and Kris Hauser. Development of a tele-nursing mobile manipulator for remote care-giving in quarantine areas. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3581–3586. IEEE, 2017.

[84] Terri Link. Guideline implementation: sterile technique. *AORN journal*, 110(4):415–425, 2019.

[85] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 21–37, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46448-0.

[86] Yu Lu, Deliang Wang, Penghe Chen, and Zhi Zhang. Design and evaluation of trustworthy knowledge tracing model for intelligent tutoring system. *IEEE Transactions on Learning Technologies*, 2024.

[87] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. Mediapipe: A framework for building perception pipelines. *ArXiv*, 2019.

[88] Karthik Mahadevan, Jonathan Chien, Noah Brown, Zhuo Xu, Carolina Parada, Fei Xia, Andy Zeng, Leila Takayama, and Dorsa Sadigh. Generative expressive robot behaviors using large language models. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 482–491, 2024.

[89] Abrar Majeedi, Ryan M McAdams, Ravneet Kaur, Shubham Gupta, Harpreet Singh, and Yin Li. Deep learning to quantify care manipulation activities in neonatal intensive care units. *npj Digital Medicine*, 7(1):172, 2024.

[90] M Marć, A Bartosiewicz, J Burzyńska, Z Chmiel, and P Januszewicz. A nursing shortage–a prospect of global and local policies. *International nursing review*, 66(1): 9–16, 2019.

[91] Sneha Mehta. Top 6 participatory design methods for your project, 2024. URL https://octet.design/journal/participatory-design-methods/.

[92] Stephanie Milani, Nicholay Topin, Manuela Veloso, and Fei Fang. Explainable reinforcement learning: A survey and comparative review. *ACM Computing Surveys*, 56 (7):1–36, 2024.

[93] Michael J Muller and Sarah Kuhn. Participatory design. *Communications of the ACM*, 36(6):24–28, 1993.

[94] Laurentiu-Marian Neagu, Eric Rigaud, Sébastien Travadel, Mihai Dascalu, and Razvan-Victor Rughinis. Intelligent tutoring systems for psychomotor training–a systematic literature review. In *International Conference on Intelligent Tutoring Systems*, pages 335–341. Springer, 2020.

[95] Heather O'Brien, Paul Cairns, and Mark Hall. A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form. *International Journal of Human-Computer Studies*, 112, 04 2018. doi: 10.1016/j.ijhcs.2018.01.004.

[96] Naomi P O'Grady. Prevention of central line-associated bloodstream infections. *The New England journal of medicine*, 2023. doi: 10.1056/NEJMra2213296.

[97] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22, 2023.

[98] André Pereira, Carlos Martinho, Iolanda Leite, and Ana Paiva. Icat, the chess player: The influence of embodiment in the enjoyment of a game. In *International Joint Conference on Autonomous Agents and Multiagent Systems*, AAMAS '08, page 1253–1256, Richland, SC, 2008. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9780981738123.

[99] Peizhu Qian and Vaibhav Unhelkar. Evaluating the role of interactivity on improving transparency in autonomous agents. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '22, page 1083–1091. International Foundation for Autonomous Agents and Multiagent Systems, 2022. ISBN 9781450392136.

[100] Peizhu Qian and Vaibhav Vasant Unhelkar. Interactively explaining robot policies to humans in integrated virtual and physical training environments. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 847–851, 2024.

[101] Carlos Quintero-Pena, Peizhu Qian, Nicole M Fontenot, Hsin-Mei Chen, Shannan K Hamlin, Lydia E Kavraki, and Vaibhav Unhelkar. Robotic tutors for nurse training: Opportunities for hri researchers. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 220–225. IEEE, 2023.

[102] Aditi Ramachandran, Chien-Ming Huang, Edward Gartland, and Brian Scassellati. Thinking aloud with a tutoring robot to enhance learning. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, pages 59–68, 2018.

[103] Aditi Ramachandran, Sarah Strohkorb Sebo, and Brian Scassellati. Personalized robot tutoring using the assistive tutor POMDP (AT-POMDP). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8050–8057, 2019.

[104] Ramya Ramakrishnan, Vaibhav Unhelkar, Ece Kamar, and Julie Shah. A bayesian approach to identifying representational errors. *arXiv preprint arXiv:2103.15171*, 2021.

[105] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016. doi: 10.1109/CVPR.2016.91.

[106] Elijah W Riddle, Divya Kewalramani, Mayur Narayan, Daniel B Jones, and Benjamin F Rush. Surgical simulation: Virtual reality to artificial intelligence. *Current Problems in Surgery*, page 101625, 2024.

[107] A Romero, J De La Hoz, and JD González. Robots in nursing education: a bibliometric analysis. In *Journal of Physics: Conference Series*, volume 1391, page 012129. IOP Publishing, 2019.

[108] Yao Rong, Tobias Leemann, Thai-Trang Nguyen, Lisa Fiedler, Peizhu Qian, Vaibhav Unhelkar, Tina Seidel, Gjergji Kasneci, and Enkelejda Kasneci. Towards human-centered explainable ai: A survey of user studies for model explanations. *IEEE transactions on pattern analysis and machine intelligence*, 2023.

[109] Fatai Sado, Chu Kiong Loo, Wei Shiung Liew, Matthias Kerzel, and Stefan Wermter. Explainable goal-driven agents and robots-a comprehensive review. *ACM Computing Surveys*, 55(10):1–41, 2023.

[110] Nicole Salomons and Brian Scassellati. Time-dependant bayesian knowledge tracing - robots that model user skills over time. *Frontiers in robotics and AI*, 2024.

[111] Nicole Salomons, Kaitlynn Taylor Pineda, Adérónké Adéjàre, and Brian Scassellati. "we make a great team!": Adults with low prior domain knowledge learn more from a peer robot than a tutor robot. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 176–184. IEEE, 2022.

[112] Sangwon Seo, Lauren R Kennedy-Metz, Marco A Zenati, Julie A Shah, Roger D Dias, and Vaibhav V Unhelkar. Towards an ai coach to infer team mental model alignment in healthcare. In *2021 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)*, pages 39–44. IEEE, 2021.

[113] S Shiny and D Venkatachalam. An analysis on the integration of ai to assist staff nurses and patients in intensive care units. In *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, pages 79–83. IEEE, 2024.

[114] Matthijs Smakman and Elly A Konijn. Robot tutors: Welcome or ethically questionable? In *Robotics in Education: Current Research and Innovations 10*, pages 376–386. Springer, 2020.

[115] Richard A Smiley, Clark Ruttinger, Carrie M Oliveira, Laura R Hudson, Richard Allgeyer, Kyrani A Reneau, Josephine H Silvestre, and Maryann Alexander. The 2020 national nursing workforce survey. *Journal of Nursing Regulation*, 12(1):S1–S96, 2021.

[116] Clay Spinuzzi. The methodology of participatory design. *Technical communication*, 52(2):163–174, 2005.

[117] Sarath Sreedharan, Tathagata Chakraborti, and Subbarao Kambhampati. Foundations of explanations as model reconciliation. *Artificial Intelligence*, 301:103558, 2021.

[118] James W Suliburk, Quentin M Buck, Chris J Pirko, Nader N Massarweh, Neal R Barshes, Hardeep Singh, and Todd K Rosengart. Analysis of human performance deficiencies associated with surgical adverse events. *JAMA network open*, 2(7):e198067–e198067, 2019.

[119] Arie Rachmad Syulistyo, Yuichiro Tanaka, and Hakaru Tamukoh. Recognizing nursing activities in endotracheal suction: Utilizing multiple readouts reservoir computing and large language models. *International Journal of*

Activity and Behavior Computing, 2024(2):1–22, 2024.
[120] Aaquib Tabrez, Shivendra Agrawal, and Bradley Hayes. Explanation-based reward coaching to improve human performance via reinforcement learning. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 249–257, 2019. doi: 10.1109/HRI.2019.8673104.
[121] Angelique Taylor, Hee Rin Lee, Alyssa Kubota, and Laurel D Riek. Coordinating clinical teams: Using robots to empower nurses to stop the line. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW): 1–30, 2019.
[122] Angelique Taylor, Tauhid Tanjim, Huajie Cao, and Hee Rin Lee. Towards collaborative crash cart robots that support clinical teamwork. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 715–724, 2024.
[123] Sam Thellman and Tom Ziemke. The perceptual belief problem: Why explainability is a tough challenge in social robotics. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(3):1–15, 2021.
[124] Andrea Thomaz. Robots in real life: Putting hri to work. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 3–3, 2023.
[125] Eylem Topbaş, Banu Terzi, Öznur Görgen, and Gülay Bingöl. Effects of different education methods in peritoneal dialysis application training on psychomotor skills and self-efficacy of nursing students. *Technology and Health Care*, 27(2):175–182, 2019.
[126] Jesus Tordesillas and Jonathan P How. Panther: Perception-aware trajectory planner in dynamic environments. *IEEE Access*, 10:22662–22677, 2022.
[127] Matthias Tschöpe, Stefan Gerd Fritsch, David Habusch, Vitor Fortes Rey, Agnes Grünerbl, and Paul Lukowicz. Evaluating deep learning models for posture and movement recognition during the abcde protocol in nurse education. In *2024 International Conference on Activity and Behavior Computing (ABC)*, pages 1–10. IEEE, 2024.
[128] Chris Varghese, Ewen M Harrison, Greg O'Grady, and Eric J Topol. Artificial intelligence in surgery. *Nature Medicine*, pages 1–12, 2024.
[129] Mudit Verma, Siddhant Bhambri, and Subbarao Kambhampati. Theory of mind abilities of large language models in human-robot interaction: An illusion? In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 36–45, 2024.
[130] David K Warren, Wasim W Quadir, Christopher S Hollenbeak, Alexis M Elward, Michael J Cox, and Victoria J Fraser. Attributable cost of catheter-associated bloodstream infections among intensive care patients in a nonteaching hospital. *Critical care medicine*, 2006.
[131] Olivia Watkins, Sandy Huang, Julius Frost, Kush Bhatia, Eric Weiner, Pieter Abbeel, Trevor Darrell, Bryan Plummer, Kate Saenko, and Anca Dragan. Explaining robot policies. *Applied AI Letters*, 2(4):e52, 2021.
[132] Katie Winkle, Emmanuel Senft, and Séverin Lemaignan. Leador: A method for end-to-end participatory design of autonomous social robots. *Frontiers in Robotics and AI*, 8:704119, 2021.
[133] H Worlikar, V Vyas Vadhiraj, Aoife Murray, J O'Connell, C Connolly, JC Walsh, and DT O'Keeffe. Is it feasible to use a humanoid robot to promote hand hygiene adherence in a hospital setting? *Infection Prevention in Practice*, page 100188, 2021.
[134] Holly A Yanco, Munjal Desai, Jill L Drury, and Aaron Steinfeld. Methods for developing trust models for intelligent systems. *Robust intelligence and trust in autonomous systems*, pages 219–254, 2016.
[135] X Jessie Yang, Vaibhav V Unhelkar, Kevin Li, and Julie A Shah. Evaluating effects of user experience and system transparency on trust in automation. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 408–416, 2017.
[136] Tuba Yilmazer and Melih Elcin. The effect of high and medium fidelity simulator in cardiopulmonary resuscitation training on nursing students' knowledge and performances. *International Journal of Caring Sciences*, 13(2):1250–1256, 2020.
[137] Rui Zeng, Yuhui Wen, Wang Zhao, and Yong-Jin Liu. View planning in robot active vision: A survey of systems, algorithms, and applications. *Computational Visual Media*, 6:225–245, 2020.
[138] Bowen Zhang and Harold Soh. Large language models as zero-shot human models for human-robot interaction. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7961–7968. IEEE, 2023.
[139] Jesse Zhang, Jiahui Zhang, Karl Pertsch, Ziyi Liu, Xiang Ren, Minsuk Chang, Shao-Hua Sun, and Joseph J Lim. Bootstrap your own skills: Learning to solve new tasks with large language model guidance. In *Conference on Robot Learning*, pages 302–325. PMLR, 2023.
[140] Matthew J Ziegler, Daniela C Pellegrini, and Nasia Safdar. Attributable mortality of central line associated bloodstream infection: systematic review and meta-analysis. *Infection*, 43:29–36, 2015.

Supplementary Material for ASTRID: A Robotic Tutor for Nurse Training to Reduce Healthcare-Associated Infections

Peizhu Qian*, Filip Bajraktari*, Carlos Quintero-Peña*, Qingxi Meng*,
Shannan Hamlin†, Lydia Kavraki‡, and Vaibhav Unhelkar‡
*Department of Computer Science, Rice University, Houston, Texas USA
pqian@rice.edu, filip.bajraktari@rice.edu, carlosq@rice.edu, qm15@rice.edu
†Center for Nursing Research, Education and Practice, Houston Methodist, Houston, Texas USA
shamlin@houstonmethodist.org
‡Ken Kennedy Institute, Rice University, Houston, Texas USA
kavraki@rice.edu, unhelkar@rice.edu

APPENDIX

a) Recruitment Methodology: We recruited participants for the Focused Interviews and Feasibility Study by sending recruitment materials to hospitals and institutions in the area. The recruitment materials were approved by Rice University IRB (IRB-FY2023-1 and IRB-FY2024-431). Participants came to Rice University for the study where we had set up a simulated training environment similar to one used in a hospital. Each session lasted $45-60$ minutes. Participants were compensated with a $25 Amazon gift card and a parking validation.

b) Tables: The next page of this appendix provides the following supplementary materials:

- Table I provides the list of questions used during the focused interviews.
- Table II provides the statements used during the feasibility study to measure ASTRID's perceived usefulness. These statements are adapted from [1]. This table reports the average score and standard deviation for each statement across participants and the overall score.
- Table III provides the statements used during the feasibility study to measure user engagement during the experiment. These statements are adapted from [2]. This table reports the average score and standard deviation for each statement across participants and the overall score.

c) Supplementary Video: Additionally, a video demonstration of ASTRID is available at http://tiny.cc/rss-2025-astrid. The video showcases ASTRID's physical interventions and human pose estimation during a practice session with a nurse participant.

REFERENCES

[1] Fred Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13:319–, 09 1989. doi: 10.2307/249008.

[2] Heather O'Brien, Paul Cairns, and Mark Hall. A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form. *International Journal of Human-Computer Studies*, 112, 04 2018. doi: 10.1016/j.ijhcs.2018.01.004.

TABLE I
INTERVIEW QUESTIONS USED IN THE FOCUSED INTERVIEWS

Section 1: Central Line Dressing Change
A. Are you familiar with the central line dressing change procedure? B. Rate level of agreement: Maintaining the sterile field during CLDC is important (1 = completely disagree to 7 = completely agree). C. Describe your approach to maintaining the sterile field. D. Rate your level of agreement: Maintaining the sterile field during CLDC is challenging. E. Describe any challenges in maintaining the sterile field. F. Describe any factors that increase the occurrence of sterile field contamination.
Section 2: Current Training Methods
A. How were you trained to perform the dressing change procedure and maintain the sterile field? B. Were the factors that increase the occurrence of sterile field contamination emphasized in the training? If yes, please elaborate. C. Did you use any technology during these training? If yes, please elaborate. D. Which methods have you used to perfect the skills required for maintaining the sterile field? E. What do you like about these methods? F. What are some limitations of these methods?
Section 3A: Brainstorming New Training Aids
A. Which features would you like to see in new training aids that can help nurses learn or perfect the skills required for maintaining the sterile field? B. How and when would nurse trainees and nurses use this aid? C. How will this aid improve nurses' ability to learn or perfect the skills required for maintaining the sterile field? D. Could there by any side-effects of using this aid? If so, please elaborate? E. Which features should not be present in this aid?
Section 3B: Feedback on a New Training Aid Prototype Prompt: We have designed and prototyped a novel training aid and would like to solicit your inputs regarding this prototype. Video demo of the prototype: http://tiny.cc/rss-2025-astrid-prelim.
A. What do you like about this prototype? B. Do you think such an aid can improve nurses' ability to learn or perfect the skills required for maintaining the sterile field? C. How and when would nurse trainees and nurses use this aid? D. What do you not like about this prototype? E. Could there be any side-effects of using such an aid? If so, please elaborate? F. Which features would you like to be added to this prototype? G. Which feature would you like to be removed from this prototype? H. How should the aid notify when a sterile field contamination occurs (audio alert, visual alert, verbally explain the reason for contamination, ...)? I. How should the aid look? J. Provide your inputs on the limitations of the system.

TABLE II
PERCEIVED USEFULNESS MEASURES (ON A 5-POINT DVAS SCALE)

	Mean	SD
1. Practicing with ASTRID would help me acquire the sterile techniques more quickly.	4.75	0.46
2. Practicing with ASTRID would improve my job performance during the central line dressing change procedure.	4.63	0.52
3. Practicing with ASTRID would improve my overall job performance.	4.38	0.52
4. Practicing with ASTRID would enhance my effectiveness on the job.	4.50	0.53
5. Practicing with ASTRID would enhance patient safety by reducing chances of healthcare-associated infections.	4.88	0.35
6. I would find ASTRID useful in nursing education.	5.00	0.00
7. I would find ASTRID useful in helping me prepare for quarterly and annual nursing evaluation.	4.75	0.46
Overall	**4.70**	**0.41**

TABLE III
USER ENGAGEMENT MEASURES (ON A 5-POINT DVAS SCALE)

	Mean	SD
1. My experience was rewarding.	4.75	0.46
2. I would recommend this system to my colleagues.	4.63	0.52
3. I would recommend this system to nursing students.	4.88	0.35
4. I was really drawn into this experience.	4.75	0.46
5. I felt involved in this experience.	5.00	0.00
6. This experience was fun.	5.00	0.00
Overall	**4.84**	**0.30**