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

***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 and educators are seeking innovative solutions to complement expert nurse educators. In response, we present a robotic tutoring system designed to aid in the training of nursing skills essential for CLABSI prevention, which is the outcome of a two-year participatory design process. 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 and motivate new avenues for research and application of human-robot interaction.**

***Index Terms*—Robotics in Healthcare, Intelligent Tutors**

I. INTRODUCTION

Registered nurses, who are with hospitalized patients around the clock, play a critical role in ensuring patient safety and delivering personalized care. As the largest segment of the hospital workforce, they are also essential to maintaining the stability of the entire healthcare system. However, this stability is now at risk due to a significant global shortage of nursing professionals [1–3]. This nurse shortage is expected to worsen due to an aging population and the long-term impact of COVID-19 [4, 5]. In the U.S. alone, for example, the healthcare sector has lost approximately half a million workers since February 2020 [6]. In addition to the growing workloads nurses face in their practice, this shortage is also 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.

Onboarding nurses is a significant undertaking that heavily depends on the availability of expert nurses. Beyond formal education in nursing schools, hospitals allocate substantial resources to train new nurses in hospital-specific practices [7]. With the increasing complexity of patient care, nursing staff must be regularly upskilled. For example, Anonymized 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 due to the shortage of expert nurses and the rise in new hires [8–10]. As a result, healthcare facilities are actively exploring innovative solutions to enhance and support nursing education [11–16].

 Author contact: {pqian, vaibhav.unhelkar}@rice.edu.
Preprint.

Figure 1. A nurse demonstrating the central line dressing change procedure on a mannekin. (top) Close up of the mannekin's arm, with the dressing.

We posit that robotic tutors designed for nursing can help address this urgent need, complementing nurse-to-nurse training. Examining this hypothesis, we present ASTRID: a robotic tutor for nursing. ASTRID is the outcome of a two-year participatory design process [17, 18] that began with requirement capture to understand nursing needs (Sec. III), included iterative system development based on these requirements (Sec. IV), and was followed by a feasibility study (Sec. V).

Depending on their specialty, registered nurses perform a wide range of complex tasks, each requiring specialized skills and training. During requirement capture, the central line dressing change [19, 20] emerged as a key nursing procedure that could benefit from robot-assisted training. Illustrated in Fig. 1, this procedure is critical for patient care and performed frequently. Moreover, it is also prone to preventable human errors that can lead to healthcare-associated infections, resulting in serious complications or, in severe cases, patient death [21–24]. These infections, however, are largely preventable through meticulous care, adherence to protocols, and regular training — currently delivered through a nurse-to-nurse model [25–27].

ASTRID aims to enhance 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

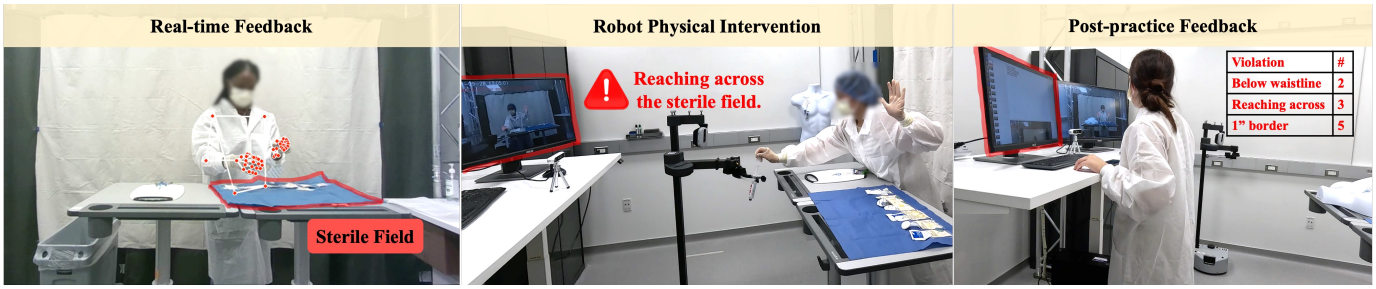

Figure 2. Nurses practicing dressing change procedures with ASTRID in a nurse training environment. ASTRID offers (left) real-time guidance, (middle) physical interventions, and (right) post-practice feedback to help nurses master "principles of sterile technique" for preventing healthcare-associated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

expert nurse. As depicted in Fig. 2, ASTRID is realized using the Stretch mobile manipulator [28], an off-board depth camera [29], and a computer. Using its perception, ASTRID monitors students as they practice dressing changes on simulated patients, providing real-time feedback when actions that could lead to infections are detected. Using its mobile manipulation, the robot simulates scenarios 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, 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.

To evaluate whether ASTRID meets its objectives, we conduct a feasibility study with nurse participants in a simulated clinical setting. Results of this study demonstrate that ASTRID can detect student errors almost as accurately as a nursing instructor, reinforcing its potential as an effective tutor. Additionally, subjective questionnaires reveal that participants find the system both useful and engaging — a critical finding for the adoption of robotic tutor systems in nursing. Finally, the evaluations suggest several promising directions for both fundamental human-robot interaction (HRI) research on robotic tutors and their practical applications in nursing education.

II. Related Work

Next, we briefly review relevant literature on nursing, robotic assistants, and intelligent tutors that informs our research.

A. Robotics for Nursing

In response to nursing shortages and growing workloads, "robotics for nursing" has become a vibrant research area [30–34]. Various robotic assistants, including commercially available products, are being developed to support nurses in homes and hospitals [35–37], with success in tasks like fetching supplies and disinfecting rooms [38–40]. While our work adopts a distinct focus by focusing on nursing education, it is informed by these robots for nursing practice. In particular, a key insight from both "robotics in nursing" research and our discussions with nurses is the importance of designing technology that does not increase their workload and involves them early on during development. Successful adoption of new technologies relies on user buy-in [41–43], and neglecting this can lead to unanticipated issues during deployment [44]. Our participatory design approach is guided by these insights, ensuring the tools we develop address real nursing needs.

B. Technology in Nursing Education

Nursing research highlights the growing need for technology in nursing education [10–13], with its role evolving rapidly since the pandemic. This shift has seen increased use of videos [45–48] and simulations [49–51]. Closer to our focus, humanoid robot-patients [13] and telepresence robots are being explored to make nursing education more accessible [14]. 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. Aligned with nursing pedagogy, which advocates for experiential learning and planned experiences [52–57], ASTRID is designed to offer student additional opportunities of "learning by doing."

C. Robotic Tutors

Intelligent tutoring systems have been developed for various learning environments, including K-12 education, corporate training, and medicine [58–64]. More recently, robotic tutors have emerged that offer additional benefits such as enhanced student interaction, improved learning outcomes, and increased trust and engagement [65–74], by leveraging their physical presence and interaction capabilities. Most robotic tutors, however, rely primarily on conversational instructions and lack mobile manipulation capabilities. HRI research highlights the value of robots that can both learn and *teach* [75], and our work seeks to bring this value to nursing education. In doing so, we also take a step towards advancing the broader field of robotic tutoring by moving beyond conversational interactions. Specifically, in addition to verbal feedback, ASTRID is designed to use its perception and mobile manipulation capabilities to assess and enhance students' physical skill execution.

III. Requirement Capture

To design the robotic tutor, we first assemble a multidisciplinary team with researchers from nursing, artificial intelligence (AI), and HRI. As introduced in Sec. I, we adopt 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

Figure 3. 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.

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: helping roboticists understand the challenges in nursing, and enabling nursing researchers to grasp the capabilities and limitations of current robotics. Roboticists also attended nurse training sessions to learn about current 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 emerged as a key focus due to it being a frequently-performed procedure, need for periodic training, and potential for robotic assistance. Before moving on to the next steps of the requirement capture process, we first provide background on this procedure.

B. Central Line Dressing Change Procedure

The placement of a central venous catheter is a commonly performed procedure in hospitals. An occlusive dressing, called the central line dressing (Fig. 1), covers the catheter's entry site to provide a protective barrier and protect patient against infections [19, 20]. Maintenance of the catheter is complex and life-critical, requiring frequent dressing changes. The supplementary material provides a detailed description of the central line dressing change procedure.

One devastating complication during dressing changes is the central line-associated bloodstream infection (CLABSI), which accounts for 17% of the almost one million healthcare-associated infections per year [24]. CLABSIs increase antibiotic exposure, hospital length of stay by 7.45 days, and risk of death among patients with CLABSI compared to those without (odds ratio, 2.75) [21–24]. In the post-COVID-19 era, unfortunately, CLABSI rates have increased [76].

Fortunately, CLABSIs are preventable with meticulous nursing care and adherence to established protocols [25–27]. When nurses use the aseptic non-touch technique, standard precautions (hand hygiene and personal protective equipment), and sterilized supplies, the risk of CLABSIs is dramatically reduced [77]. These safety protocols, referred to as the "principles of sterile technique," outline essential rules for maintaining sterility during dressing changes. Four key rules, which we focus on in this paper, are illustrated in Fig. 3.

Adherence to the principles of sterile technique is challenging, with failure rates reaching 45% or higher [78]. Nurses often break associated rules inadvertently, especially under rising workloads, due to increasing nursing burnout, or in challenging situations caused by interruptions and disruptions. While periodic skills validation on maintaining a sterile field can help improve compliance, sustaining this within the current nurse-to-nurse training model is difficult. For instance, at Anonymized hospital, 15+ experienced nurses are needed daily to facilitate such training, straining the already limited workforce and posing patient safety risks. Furthermore, the shortage of hospital-based nursing instructors limits training opportunities and innovative solutions are urgently needed [8, 19, 26]. Informed by this need, we explore the potential of robotic tutors for CLABSI-prevention education and training.

C. 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.

1) Methodology: The interview protocol was approved by Anonymized IRB and structured in three parts, each addressing a specific goal. The first part focused on understanding participants' experiences with dressing changes 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 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.

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) Finding #1: Nursing students find maintaining a sterile field challenging. All participants agreed that maintaining sterility is crucial during a dressing change. While it may

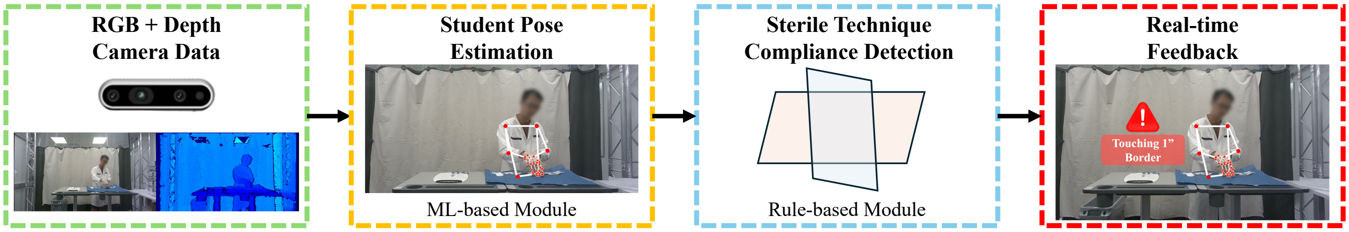

Figure 4. Overview of ASTRID's system architecture for providing real-time guidance regarding the sterile technique to nursing students.

seem straightforward in controlled settings, real-world factors complicate the task. Nurses often face scenarios such as interruptions from patients or family members, children running in hospital hallways, patients moving unexpectedly, or unforeseen medical complications. These scenarios may not always be covered in nursing schools, requiring supplemental on-the-job training. Participants from all groups noted that periodic training addressing real-world challenges is beneficial for nurses at all experience levels.

Finding #2: Nurses prefer training aids that are accessible and encouraging. New nurses and interns can struggle to find opportunities to practice their skills. Participants noted that new nurses undergo a one-week orientation, which includes a class on the dressing change procedure; however, with only one educator for $30 - 40$ students, personalized attention can be limited. Some students feel overwhelmed by the amount of information and find it difficult to secure opportunities for follow-up practice with experienced nurses. Even when they have the chance to practice with an experienced nurse, some new nurses may feel anxious and nervous. All participants agree that a training aid that is available and encouraging can be beneficial. These features should be reflected in the design of a robotic tutor, which should provide nursing students additional opportunities to practice their skills in an encouraging manner.

D. System Requirements

Based on the exploratory phase, nursing research, and interviews, we distill key requirements (**Rx**) for a robot that assists in CLABSI-prevention training. The robotic tutor must:

R1. detect compliance with the sterile technique;
R2. provide real-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;
R6. be perceived as encouraging by nursing students; and
R7. be perceived as engaging by nursing students.

IV. Prototype Design

Guided by the system requirements, we design ASTRID: the *Automated Sterile Technique Review and Instruction Device*. As shown in Fig. 2, ASTRID is a robotic tutor prototype realized using the Stretch mobile manipulator [28], an off-board depth camera [29], and a computer. Through the software architecture outlined in Fig. 4, the system monitors nursing students as they practice the central line dressing change procedure, provides real-time feedback on their compliance with the sterile technique, simulates error-prone scenarios like disruptions, and generates a summary report for performance review and improvement. It is intended for use in training environments (Fig. 5), and not patient rooms or hospital floors. In this section, we detail its key features and their implementation.

A. Detecting Student Compliance with the Sterile Technique

Illustrated in Fig. 3, ASTRID considers four key principles of sterile technique, which require that once the sterile field is established, the nurse should:

- keep hands above their waistline;
- not turn back away from the sterile field;
- not reach across the sterile field; and
- not touch the 1-inch border of the sterile field.

As nursing students practice the central line dressing change procedure, ASTRID monitors their task execution through its depth camera and detects compliance with these four principles via computer vision algorithms.

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 in side pouches, making detection complex. To ensure robust detection, ASTRID utilizes a manual calibration method.[1] Since the sterile field remains static during the procedure, calibration is required only once. In particular, before training begins, the user is shown a real-time image of the training scenario on the monitor and marks the drape's corners using a mouse. We find that this method ensures accuracy across different environments and user conditions, such as varied backgrounds or user heights. 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, an open-source framework designed for real-time machine learning applications [79]. 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. We set the visibility (a hyperparameter of MediaPipe)

[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.

of irrelevant landmarks to 0 to enhance robustness of this automated pose estimation. This hyperparameter tuning is critical as 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.

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. First, ASTRID computes the 3D coordinates of each landmark using the pinhole camera model. Then, to partition the 3D space into sterile and non-sterile regions, it calculates 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 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.

that violate the four principles of sterile technique, respectively.

B. Providing Feedback to Students

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

1) Real-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)." During requirement capture, participants were asked about their preferences for alerts: visual, short audio beeps, constant long beeps, or descriptive audio. While most preferred detailed audio alerts, they also cautioned about "alarm fatigue," where excessive alerts reduce attentiveness. As a result, we decided to use both visual and audio alerts that convey the non-compliant behavior.

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 computer (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 key frames and associated warnings.

Figure 5. Artificial rendering of the training environment shown in Fig. 2, highlighting its key elements: nurse, simulated patient, medical supplies, and ASTRID (composed of the robot, depth camera, and computer screen).

C. Simulating Challenging Nursing Scenarios

We design ASTRID to offer three levels of training – novice, intermediate, and advanced – which increase in complexity, simulating challenging scenarios that nurses may encounter during dressing change procedures.

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.

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. Our prototype, currently, offers two such scenarios:

- *(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. 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. In such cases, experienced nurses would typically inform the patient that they will retrieve the item after the procedure.

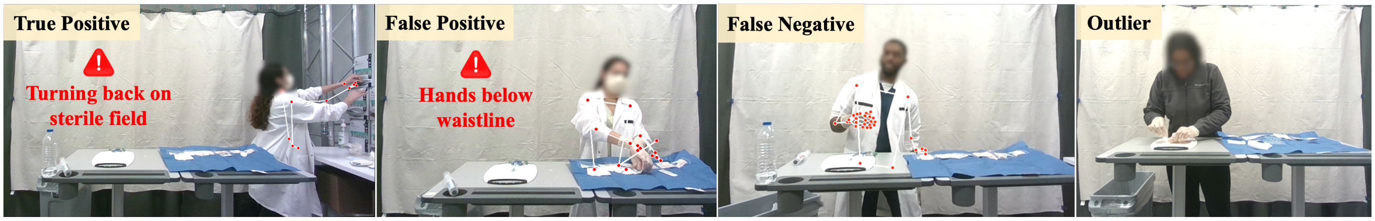

Figure 6. Examples of ASTRID's compliance detection from the feasibility study: (True Positive) ASTRID correctly detects the nurse turning their back to the sterile field; (False Positive) ASTRID mistakenly identifies the hands as below the waistline due to occlusion; (False Negative) ASTRID misses the 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.

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.

V. FEASIBILITY STUDY

We conducted a feasibility study to evaluate ASTRID.[2] This IRB-approved study involved nine participants – two experienced nurses from the requirement capture interviews and seven recent graduates or nurses with two years or less of experience. The experiment was held in the training environment shown in Fig. 2, an artificial rendering of which is depicted in Fig. 5.

A. Materials

The experiment site was set up to mimic a central line dressing change scenario, 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. ASTRID was realized using the Stretch robot, an Intel RealSense D455i depth camera, and a computer monitor. 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 the central line dressing change procedure 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 central line dressing change 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-visuals 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-C3.
- **Test 3** 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**–**R7**) 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) [80], 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 the ASTRID detected a violation but the expert did not. Fig. 6 provides additional examples.

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

Table I
PERCEIVED USEFULNESS MEASURES (ON A 5-POINT LIKERT 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

Table II
USER ENGAGEMENT MEASURES (ON A 5-POINT LIKERT 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

Table III
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)

*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. 7). Participants rated each feature on a 5-point Likert scale, from *not useful at all (1)* to *extremely useful (5)*.

*3) Measures for **R5–R7**:* The post-experiment survey included two additional sections: one assessing the perceived usefulness of ASTRID in nursing education and the other evaluating user engagement while interacting with ASTRID. Perceived usefulness was measured using statements shown in Table I, designed based on a widely used scale for predicting user acceptance of technology [81]. Participants rated these statements on a 5-point Likert scale, ranging from *extremely unlikely (1)* to *extremely likely (5)*. Engagement was assessed using statements listed in Table II, adapted from the User Engagement Scale [82]. This scale also utilized a 5-point Likert scale: *strongly disagree (1)* to *strongly agree (5)*.

D. Findings

Nine nurses participated in the feasibility study. Data from one participant ($P2$) was excluded from analysis as an outlier due to issues with the table height and camera configuration. Although we had opted for a height-adjustable table to accommodate participants of different height, it was still too low for this participant. 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 of the table height-related issue, 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.

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 III summarizes related metrics, which also suggest that ASTRID satisfies requirement **R1**.

2) Participants perceive ASTRID *as highly useful for nursing education:* When asked for their level of agreement with the statement "I would find ASTRID useful in nursing education," all participants rate it as *extremely likely*. For another statement, "Practicing with ASTRID would enhance patient safety by reducing chances of healthcare-associated infections," the average rating was 4.88. Table I summarizes responses to the perceived usefulness measures, with all statements receiving a high average score. Fig. 7 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.

3) Participants perceive ASTRID *as engaging and supportive:* Participants' responses to the user engagement statements, summarized in Table II, indicate that they found their interaction with ASTRID to be highly engaging. All statements received an average score of 4.63 or above on a 5-point Likert scale. Moreover, 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

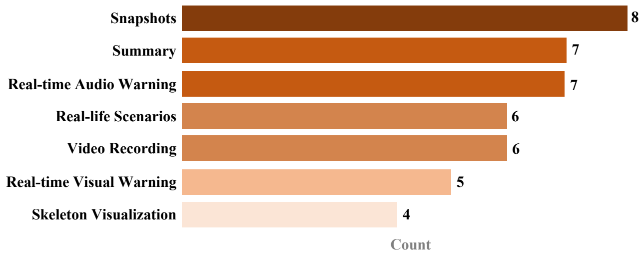

Figure 7. Perceived usefulness of ASTRID's individual features: the figure shows the number of participants that rated a feature to be *extremely useful*.

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**.

VI. Conclusion

We conclude by discussing the implications of our findings for both nursing education and human-robot interaction.

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 [8–10]. Further, our approach reaffirms the importance of involving nurses early on during the design and development of new technology.

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

- ASTRID addresses four principles of the sterile technique; however, this technique is more comprehensive [25–27]. 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 [83], and high-sterility environments like operating rooms (OR) [84–87].
- 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. 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 Human-Robot Interaction

1) Key Contributions: Our work introduces a new domain for HRI: nursing education, a field in need of transformative solutions. We developed ASTRID, a robotic tutor for nursing and demonstrated its potential. Our iterative, stakeholder-driven approach reaffirmed the value of participatory design. Moreover, unlike existing robotic tutors that primarily rely on conversational instructions, ASTRID also incorporates physical interventions through its mobile manipulation. We see this as a small but important step toward advancing the field of robotic tutoring beyond conversational or screen-based interactions.

2) Limitations and Future Work: Our investigation also reveals directions for HRI research, both for enhancing nursing-specific tutors and advancing the broader field of robotic tutors:

- During evaluations, ASTRID struggled with taller participants (e.g., Fig. 6-outlier), 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 [88–93] and mobile perception [94–97].
- 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 [98–103]. 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.
- Additionally, as robotic tutors become more capable, trust calibration is a critical concern to prevent students from over-relying on these systems [104–109]. This is especially important for robotic tutors, as their teaching role may lead students to inherent trust them more than robotic assistants or peers [110–115]. One way to address this is by explaining the robot's capabilities and limitations to students and involving instructors in the process [116–133]. 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 [134–138].

Lastly, while we outline future research for nursing and HRI separately, we emphasize the need for cross-disciplinary collaboration. Our research reaffirms that developing HRI systems requires both inputs from domain experts and use of HRI methods. This integrated approach is key to developing safe, responsible, and sustainable human-centered technology.

References

[1] M. Marć, A. Bartosiewicz, J. Burzyńska, Z. Chmiel, and P. Januszewicz. "A nursing shortage–a prospect of global and local policies". In: *International nursing review* 66.1 (2019), pp. 9–16.

[2] R. A. Smiley, C. Ruttinger, C. M. Oliveira, L. R. Hudson, R. Allgeyer, K. A. Reneau, J. H. Silvestre, and M. Alexander. "The 2020 national nursing workforce survey". In: *Journal of Nursing Regulation* 12.1 (2021), S1–S96.

[3] L. M. Haddad, P. Annamaraju, and T. J. Toney-Butler. "Nursing shortage". In: *StatPearls [Internet]* (2020).

[4] Bureau of Labor Statistics, U.S. Department of Labor. *Occupational Outlook Handbook, Registered Nurses*.

[5] T. Bateman, S. Hobaugh, E. Pridgen, and A. Reddy. *Demand for healthcare workers will outpace supply by 2025*. Mercer, 2021.

[6] E. Wager, K. Amin, C. Cox, and P. Hughes-Cromwick. *What impact has the coronavirus pandemic had on health employment?* Peterson-KFF Health System Tracker, 2021.

[7] M. Ito, H. Mitsunaga, and T. Ibe. "Survey of Onboarding Programs of Hospitals for Newly Hired Experienced Nurses." In: *Journal of St. Luke's Society for Nursing Research (SLNR)* 24 (2021).

[8] J. Fitzwater, J. McNeill, D. Monsivais, and F. Nunez. "Using Simulation to Facilitate Transition to the Nurse Educator Role: An Integrative Review". In: *Nurse educator* (2021).

[9] C. Quintero-Pena, P. Qian, N. M. Fontenot, H.-M. Chen, S. K. Hamlin, L. E. Kavraki, and V. Unhelkar. "Robotic Tutors for Nurse Training: Opportunities for HRI Researchers". In: *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE. 2023, pp. 220–225.

[10] K. M. Huun and J. E. Slaven. "Robotic Telepresence and Face-to-Face Collaborative Nursing Simulation: A Correlational, Cross-Sectional Study". In: *Clinical Simulation in Nursing* 90 (2024), p. 101525.

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

[12] A. Dante, A. Marcotullio, V. Masotta, V. Caponnetto, C. La Cerra, L. Bertocchi, C. Petrucci, and C. 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*. Springer. 2021, pp. 111–118.

[13] A. Dante, C. La Cerra, V. Masotta, V. Caponnetto, L. Bertocchi, A. Marcotullio, F. Ferraiuolo, C. M. Alfes, and C. 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*. Springer. 2022, pp. 217–226.

[14] A. Abuatiq, R. Brown, C. Plemmons, B. Walstrom, C. Hultman, D. Currier, M. Schmit, V. Kvigne, L. Horsley, and H. Mennenga. "Nursing faculty and students' satisfaction with telepresence robots during the COVID-19 pandemic". In: *Nurse Educator* 47.2 (2022), E39–E42.

[15] J. Dhaussy, L. Kemken, M.-T. Pugliese, A. Forestier, and S. Boloré. "Using simulation to adapt nursing education to times of crisis: A scoping review during Covid-19 pandemic". In: *Teaching and Learning in Nursing* (2024).

[16] 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)*. IEEE. 2024, pp. 79–83.

[17] M. J. Muller and S. Kuhn. "Participatory design". In: *Communications of the ACM* 36.6 (1993), pp. 24–28.

[18] C. Spinuzzi. "The methodology of participatory design". In: *Technical communication* 52.2 (2005), pp. 163–174.

[19] L. A. Gorski, L. Hadaway, M. E. Hagle, D. Broadhurst, S. Clare, T. Kleidon, B. M. Meyer, B. Nickel, S. Rowley, E. Sharpe, and M. Alexander. "Infusion Therapy Standards of Practice, 8th Edition". In: *Journal of infusion nursing: the official publication of the Infusion Nurses Society* (2021).

[20] M. A. Hicks, P. Popowicz, and P. P. Lopez. "Central Line Management". In: *StatPearls [Internet]*. StatPearls Publishing, 2023.

[21] D. K. Warren, W. W. Quadir, C. S. Hollenbeak, A. M. Elward, M. J. Cox, and V. J. Fraser. "Attributable cost of catheter-associated bloodstream infections among intensive care patients in a nonteaching hospital". In: *Critical care medicine* (2006).

[22] M. J. Ziegler, D. C. Pellegrini, and N. Safdar. "Attributable mortality of central line associated bloodstream infection: systematic review and meta-analysis". In: *Infection* 43 (2015), pp. 29–36.

[23] T. Bysshe, Y. Gao, K. Heaney-Huls, J. Hockenberry, L. Hovey, A. M. Laffan, S. Lee, D. J. Murphy, and E. Watts. *Estimating the Additional Hospital Inpatient Cost and Mortality Associated With Selected Hospital-Acquired Conditions*. Tech. rep. Agency for Healthcare Research and Quality, 2017.

[24] J. D. Forrester, P. M. Maggio, and L. Tennakoon. "Cost of Health Care-Associated Infections in the United States". In: *Journal of patient safety* (2022).

[25] T. Link. "Guideline implementation: sterile technique". In: *AORN journal* 110.4 (2019), pp. 415–425.

[26] N. Buetti, J. Marschall, M. Drees, M. G. Fakih, L. Hadaway, L. L. Maragakis, E. Monsees, S. Novosad, N. P. O'Grady, M. E. Rupp, J. Wolf, D. Yokoe, and L. A. Mermel. "Strategies to prevent central line-associated bloodstream infections in acute-care hospitals: 2022 Update". In: *Infection control and hospital epidemiology* (2022).

[27] N. P. O'Grady. "Prevention of Central Line-Associated Bloodstream Infections". In: *The New England journal of medicine* (2023).

[28] C. C. Kemp, A. Edsinger, H. M. Clever, and B. Matulevich. "The design of stretch: A compact, lightweight mobile manipulator for indoor human environments". In: *2022 International Conference on Robotics and Automation (ICRA)*. IEEE. 2022, pp. 3150–3157.

[29] Intel. *Intel RealSense Technology*. 2024.

[30] T. E. Kirschling, S. S. Rough, and B. C. Ludwig. "Determining the feasibility of robotic courier medication delivery in a hospital setting". In: *American Journal of Health-System Pharmacy* 66.19 (2009), pp. 1754–1762.

[31] R. Bloss. "Mobile hospital robots cure numerous logistic needs". In: *Industrial Robot: An International Journal* (2011).

[32] E. Ackerman. "Diligent Robotics Bringing Autonomous Mobile Manipulation to Hospitals". In: *IEEE Spectrum* (2018).

[33] M. Kangasniemi, S. Karki, N. Colley, and A. Voutilainen. "The use of robots and other automated devices in nurses' work: An integrative review". In: *International journal of nursing practice* 25.4 (2019), e12739.

[34] E. Ackerman. "Akara Robotics Turns TurtleBot Into Autonomous UV Disinfecting Robot". In: *IEEE Spectrum* (2020).

[35] M. Gombolay, X. J. Yang, B. Hayes, N. Seo, Z. Liu, S. Wadhwania, T. Yu, N. Shah, T. Golen, and J. Shah. "Robotic assistance in the coordination of patient care". In: *The International Journal of Robotics Research* 37.10 (2018), pp. 1300–1316.

[36] A. Taylor, H. R. Lee, A. Kubota, and L. D. Riek. "Coordinating clinical teams: Using robots to empower nurses to stop the line". In: *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW (2019), pp. 1–30.

[37] A. Thomaz. "Robots in Real Life: Putting HRI to Work". In: *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 2023, pp. 3–3.

[38] Z. Li, P. Moran, Q. Dong, R. J. Shaw, and K. 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)*. IEEE. 2017, pp. 3581–3586.

[39] S. Abubakar, S. K. Das, C. Robinson, M. N. Saadatzi, M. C. Logsdon, H. Mitchell, D. Chlebowy, and D. 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)*. IEEE. 2020, pp. 1408–1414.

[40] H. Worlikar, V. V. Vadhiraj, A. Murray, J. O'Connell, C. Connolly, J. Walsh, and D. O'Keeffe. "Is it feasible to use a humanoid robot to promote hand hygiene adherence in a hospital setting?" In: *Infection Prevention in Practice* (2021), p. 100188.

[41] R. Parasuraman and V. Riley. “Humans and automation: Use, misuse, disuse, abuse”. In: *Human factors* 39.2 (1997), pp. 230–253.
[42] M. Venus, D. Stam, and D. Van Knippenberg. “Visions of change as visions of continuity”. In: *Academy of Management Journal* 62.3 (2019), pp. 667–690.
[43] Y. Zhou, Z. Li, and Y. Li. “Interdisciplinary collaboration between nursing and engineering in health care: A scoping review”. In: *International journal of nursing studies* 117 (2021), p. 103900.
[44] M. Tietze and S. McBride. *Robotics and the Impact on Nursing Practice*. 2020.
[45] G. A. Dinndorf-Hogenson, C. Hoover, J. L. Berndt, B. Tollefson, J. Peterson, and N. Laudenbach. “Applying the flipped classroom model to psychomotor skill acquisition in nursing”. In: *Nursing education perspectives* 40.2 (2019), pp. 99–101.
[46] N. S. Ali and B. John. “Examining the efficacy of online self-paced interactive video-recordings in nursing skill competency learning: seeking preliminary evidence through an action research”. In: *Medical Science Educator* 29 (2019), pp. 463–473.
[47] Y.-H. Chuang, F.-C. Lai, C.-C. Chang, and H.-T. Wan. “Effects of a skill demonstration video delivered by smartphone on facilitating nursing students’ skill competencies and self-confidence: A randomized controlled trial study”. In: *Nurse education today* 66 (2018), pp. 63–68.
[48] B. Devi, B. Khandelwal, and M. Das. “Comparison of the effectiveness of video-assisted teaching program and traditional demonstration on nursing students learning skills of performing obstetrical palpation”. In: *Iranian journal of nursing and midwifery research* 24.2 (2019), p. 118.
[49] J. Chen, J. Yang, F. Hu, S.-H. Yu, B.-X. Yang, Q. Liu, and X.-P. Zhu. “Standardised simulation-based emergency and intensive care nursing curriculum to improve nursing students’ performance during simulated resuscitation: a quasi-experimental study”. In: *Intensive and Critical Care Nursing* 46 (2018), pp. 51–56.
[50] E. Topbaş, B. Terzi, Ö. Görgen, and G. Bingöl. “Effects of different education methods in peritoneal dialysis application training on psychomotor skills and self-efficacy of nursing students”. In: *Technology and Health Care* 27.2 (2019), pp. 175–182.
[51] T. Yilmazer and M. Elcin. “The Effect of High and Medium Fidelity Simulator in Cardiopulmonary Resuscitation Training on Nursing Students’ Knowledge and Performances”. In: *International Journal of Caring Sciences* 13.2 (2020), pp. 1250–1256.
[52] J. Kaakinen and E. Arwood. “Systematic review of nursing simulation literature for use of learning theory”. In: *International Journal of Nursing Education Scholarship* 6.1 (2009).
[53] B. K. Robinson and V. Dearmon. “Evidence-based nursing education: Effective use of instructional design and simulated learning environments to enhance knowledge transfer in undergraduate nursing students”. In: *Journal of Professional Nursing* 29.4 (2013), pp. 203–209.
[54] P. Lavoie, C. Michaud, M. Belisle, L. Boyer, E. Gosselin, M. Grondin, C. Larue, S. Lavoie, and J. Pepin. “Learning theories and tools for the assessment of core nursing competencies in simulation: A theoretical review”. In: *Journal of Advanced Nursing* 74.2 (2018), pp. 239–250.
[55] D. Rim and H. Shin. “Effective instructional design template for virtual simulations in nursing education”. In: *Nurse Education Today* 96 (2021), p. 104624.
[56] M. C. Hehman, N. M. Fontenot, G. K. Drake, and R. S. Musgrove. “Leveraging digital technology in nursing”. In: *Health Emergency and Disaster Nursing* 10.1 (2023), pp. 41–45.
[57] S. Hamlin, N. Fontenot, S. Hooker, and H. Chen. “Systems-Based Physical Assessments: Earlier Detection of Patient Deterioration and Reduced Mortality.” In: *American Journal of Critical Care* In Press (2023).
[58] A. Ramachandran, S. S. Sebo, and B. Scassellati. “Personalized robot tutoring using the assistive tutor POMDP (AT-POMDP)”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 2019, pp. 8050–8057.
[59] R. Ramakrishnan, V. Unhelkar, E. Kamar, and J. Shah. “A Bayesian Approach to Identifying Representational Errors”. In: *arXiv preprint arXiv:2103.15171* (2021).
[60] L.-M. Neagu, E. Rigaud, S. Travadel, M. Dascalu, and R.-V. Rughinis. “Intelligent tutoring systems for psychomotor training–a systematic literature review”. In: *International Conference on Intelligent Tutoring Systems*. Springer. 2020, pp. 335–341.
[61] 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*. Vol. 4. IEEE. 2001, pp. 4024–4027.
[62] M. Hospers, E. Kroezen, A. Nijholt, R. op den Akker, and D. Heylen. “An agent-based intelligent tutoring system for nurse education”. In: *Applications of Software Agent Technology in the Health Care Domain* (2003), pp. 143–159.
[63] M. Abuazizeh, T. Kirste, and K. 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*. 2020, pp. 1–7.
[64] M. Abuazizeh, K. Yordanova, and T. Kirste. “Affect-Aware Conversational Agent for Intelligent Tutoring of Students in Nursing Subjects”. In: *Intelligent Tutoring Systems*. Ed. by A. I. Cristea and C. Troussas. Cham: Springer International Publishing, 2021, pp. 497–502.
[65] D. Leyzberg, S. Spaulding, M. Toneva, and B. Scassellati. “The physical presence of a robot tutor increases cognitive learning gains”. In: *Proceedings of the annual meeting of the cognitive science society*. Vol. 34. 2012.
[66] W. A. Bainbridge, J. W. Hart, E. S. Kim, and B. Scassellati. “The benefits of interactions with physically present robots over video-displayed agents”. In: *International Journal of Social Robotics* 3 (2011), pp. 41–52.
[67] A. Pereira, C. Martinho, I. Leite, and A. 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. Estoril, Portugal: International Foundation for Autonomous Agents and Multiagent Systems, 2008, pp. 1253–1256.
[68] P. Alves-Oliveira, T. Ribeiro, S. Petisca, E. Di Tullio, F. S. Melo, and A. 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*. Springer. 2015, pp. 21–30.
[69] G. Castellano, A. Paiva, A. Kappas, R. Aylett, H. Hastie, W. Barendregt, F. Nabais, and S. 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*. Springer. 2013, pp. 733–736.
[70] P. Alves-Oliveira, S. Janarthanam, A. Candeias, A. Deshmukh, T. Ribeiro, H. Hastie, A. Paiva, and R. Aylett. “Towards dialogue dimensions for a robotic tutor in collaborative learning scenarios”. In: *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE. 2014, pp. 862–867.
[71] P. Alves-Oliveira, P. Sequeira, and A. Paiva. “The role that an educational robot plays”. In: *2016 25th IEEE International symposium on robot and human interactive communication (RO-MAN)*. IEEE. 2016, pp. 817–822.
[72] M. Donnermann, P. Schaper, and B. Lugrin. “Social robots in applied settings: A long-term study on adaptive robotic tutors in higher education”. In: *Frontiers in Robotics and AI* 9 (2022), p. 831633.
[73] A. Ramachandran, C.-M. Huang, E. Gartland, and B. Scassellati. “Thinking aloud with a tutoring robot to enhance learning”. In: *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*. 2018, pp. 59–68.
[74] D. Leyzberg, A. Ramachandran, and B. Scassellati. “The effect of personalization in longer-term robot tutoring”. In: *ACM Transactions on Human-Robot Interaction (THRI)* 7.3 (2018), pp. 1–19.
[75] T. Adamson, D. Ghose, S. C. Yasuda, L. J. S. Shepard, M. A. Lewkowicz, J. Duan, and B. Scassellati. “Why we should build robots that both teach and learn”. In: *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*. 2021, pp. 187–196.
[76] J. LeRose, A. Sandhu, J. Polistico, J. Ellsworth, M. Cranis, L. Jabbo, L. Cullen, J. Moshos, L. Samavati, and T. Chopra. “The impact of coronavirus disease 2019 (COVID-19) response on central-line-associated bloodstream infections and blood culture contamination rates at a tertiary-care center in the Greater Detroit area”. In: *Infection control and hospital epidemiology* (2021).
[77] D. C. V. Reyes, M. Bloomer, and J. Morphet. “Prevention of central venous line associated bloodstream infections in adult intensive care units: A systematic review”. In: *Intensive and Critical Care Nursing* 43 (2017), pp. 12–22.

[78] I. Shapey, M. Foster, T. Whitehouse, P. Jumaa, and J. Bion. "Central venous catheter-related bloodstream infections: improving post-insertion catheter care". In: *Journal of Hospital Infection* 71.2 (2009), pp. 117–122.

[79] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann. "MediaPipe: A Framework for Building Perception Pipelines". In: *ArXiv* (2019).

[80] O. P. Friard, M. Gamba, et al. "Behavioral Observation Research Interactive Software (BORIS)". In: (2016).

[81] F. Davis and F. Davis. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology". In: *MIS Quarterly* 13 (Sept. 1989), pp. 319–.

[82] H. O'Brien, P. Cairns, and M. Hall. "A Practical Approach to Measuring User Engagement with the Refined User Engagement Scale (UES) and New UES Short Form". In: *International Journal of Human-Computer Studies* 112 (Apr. 2018).

[83] A. Bianchi, S. W. Leslie, and G. T. Chesnut. "Difficult foley catheterization". In: *StatPearls [Internet]* (2023).

[84] J. W. Suliburk, Q. M. Buck, C. J. Pirko, N. N. Massarweh, N. R. Barshes, H. Singh, and T. K. Rosengart. "Analysis of human performance deficiencies associated with surgical adverse events". In: *JAMA network open* 2.7 (2019), e198067–e198067.

[85] S. Seo, L. R. Kennedy-Metz, M. A. Zenati, J. A. Shah, R. D. Dias, and V. 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)*. IEEE. 2021, pp. 39–44.

[86] E. W. Riddle, D. Kewalramani, M. Narayan, D. B. Jones, and B. F. Rush. "Surgical Simulation: Virtual Reality to Artificial Intelligence". In: *Current Problems in Surgery* (2024), p. 101625.

[87] C. Varghese, E. M. Harrison, G. O'Grady, and E. J. Topol. "Artificial intelligence in surgery". In: *Nature Medicine* (2024), pp. 1–12.

[88] W. Choi, C. Pantofaru, and S. 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)*. IEEE. 2011, pp. 1076–1083.

[89] C. Gu, C. Sun, D. A. Ross, C. Vondrick, C. Pantofaru, Y. Li, S. Vijayanarasimhan, G. Toderici, S. Ricco, R. 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*. 2018, pp. 6047–6056.

[90] A. R. Syulistyo, Y. Tanaka, and H. Tamukoh. "Recognizing Nursing Activities in Endotracheal Suction: Utilizing Multiple Readouts Reservoir Computing and Large Language Models". In: *International Journal of Activity and Behavior Computing* 2024.2 (2024), pp. 1–22.

[91] M. Hu, L. Wang, S. Yan, D. Ma, Q. Ren, P. Xia, W. Feng, P. Duan, L. Ju, and Z. Ge. "Nurvid: A large expert-level video database for nursing procedure activity understanding". In: *Advances in Neural Information Processing Systems* 36 (2024).

[92] M. Tschöpe, S. G. Fritsch, D. Habusch, V. F. Rey, A. Grünerbl, and P. 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)*. IEEE. 2024, pp. 1–10.

[93] A. Majeedi, R. M. McAdams, R. Kaur, S. Gupta, H. Singh, and Y. Li. "Deep learning to quantify care manipulation activities in neonatal intensive care units". In: *npj Digital Medicine* 7.1 (2024), p. 172.

[94] D. Falanga, P. Foehn, P. Lu, and D. Scaramuzza. "PAMPC: Perception-aware model predictive control for quadrotors". In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE. 2018, pp. 1–8.

[95] B. Ichter, B. Landry, E. Schmerling, and M. Pavone. "Perception-aware motion planning via multiobjective search on gpus". In: *Robotics Research: The 18th International Symposium ISRR*. Springer. 2020, pp. 895–912.

[96] R. Zeng, Y. Wen, W. Zhao, and Y.-J. Liu. "View planning in robot active vision: A survey of systems, algorithms, and applications". In: *Computational Visual Media* 6 (2020), pp. 225–245.

[97] J. Tordesillas and J. P. How. "PANTHER: Perception-aware trajectory planner in dynamic environments". In: *IEEE Access* 10 (2022), pp. 22662–22677.

[98] E. M. Bender, T. Gebru, A. McMillan-Major, and S. 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*. 2021, pp. 610–623.

[99] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein. "Generative agents: Interactive simulacra of human behavior". In: *Proceedings of the 36th annual acm symposium on user interface software and technology*. 2023, pp. 1–22.

[100] B. Zhang and H. 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)*. IEEE. 2023, pp. 7961–7968.

[101] A. Bouzida, A. Kubota, D. Cruz-Sandoval, E. W. Twamley, and L. 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*. 2024, pp. 55–64.

[102] C. Y. Kim, C. P. Lee, and B. Mutlu. "Understanding large-language model (llm)-powered human-robot interaction". In: *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 2024, pp. 371–380.

[103] M. Verma, S. Bhambri, and S. 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*. 2024, pp. 36–45.

[104] M. Desai, P. Kaniarasu, M. Medvedev, A. Steinfeld, and H. Yanco. "Impact of robot failures and feedback on real-time trust". In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE. 2013, pp. 251–258.

[105] H. A. Yanco, M. Desai, J. L. Drury, and A. Steinfeld. "Methods for developing trust models for intelligent systems". In: *Robust intelligence and trust in autonomous systems* (2016), pp. 219–254.

[106] X. J. Yang, V. V. Unhelkar, K. Li, and J. 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*. 2017, pp. 408–416.

[107] M. Lewis, K. Sycara, and P. Walker. "The role of trust in human-robot interaction". In: *Foundations of trusted autonomy* (2018), pp. 135–159.

[108] B. C. Kok and H. Soh. "Trust in robots: Challenges and opportunities". In: *Current Robotics Reports* 1.4 (2020), pp. 297–309.

[109] Y. Guo and X. J. Yang. "Modeling and predicting trust dynamics in human–robot teaming: A Bayesian inference approach". In: *International Journal of Social Robotics* 13.8 (2021), pp. 1899–1909.

[110] C. Birmingham, Z. Hu, K. Mahajan, E. Reber, and M. 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)*. IEEE. 2020, pp. 8019–8026.

[111] K. A. Brink and H. M. Wellman. "Robot teachers for children? Young children trust robots depending on their perceived accuracy and agency." In: *Developmental Psychology* 56.7 (2020), p. 1268.

[112] M. Smakman and E. A. Konijn. "Robot tutors: Welcome or ethically questionable?" In: *Robotics in Education: Current Research and Innovations 10*. Springer. 2020, pp. 376–386.

[113] C. Conati, O. Barral, V. Putnam, and L. Rieger. "Toward personalized XAI: A case study in intelligent tutoring systems". In: *Artificial intelligence* 298 (2021), p. 103503.

[114] Y. Lu, D. Wang, P. Chen, and Z. Zhang. "Design and Evaluation of Trustworthy Knowledge Tracing Model for Intelligent Tutoring System". In: *IEEE Transactions on Learning Technologies* (2024).

[115] N. Salomons, K. T. Pineda, A. Adéjàre, and B. 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)*. IEEE. 2022, pp. 176–184.

[116] B. Hayes and J. A. Shah. "Improving robot controller transparency through autonomous policy explanation". In: *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*. 2017, pp. 303–312.

[117] D. Amir and O. Amir. "Highlights: Summarizing agent behavior to people". In: *Proceedings of the 17th international conference on autonomous agents and multiagent systems*. 2018, pp. 1168–1176.

[118] M. Kwon, S. H. Huang, and A. D. Dragan. "Expressing robot incapability". In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. 2018, pp. 87–95.

[119] O. Amir, F. Doshi-Velez, and D. Sarne. "Summarizing agent strategies". In: *Autonomous Agents and Multi-Agent Systems* 33 (2019), pp. 628–644.

[120] S. Anjomshoae, A. Najjar, D. Calvaresi, and K. 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*. International Foundation for Autonomous Agents and Multiagent Systems. 2019, pp. 1078–1088.

[121] T. Chakraborti, A. Kulkarni, S. Sreedharan, D. E. Smith, and S. Kambhampati. "Explicability? legibility? predictability? transparency? privacy? security? the emerging landscape of interpretable agent behavior". In: *Proceedings of the international conference on automated planning and scheduling*. Vol. 29. 2019, pp. 86–96.

[122] I. Lage, D. Lifschitz, F. Doshi-Velez, and O. Amir. "Exploring computational user models for agent policy summarization". In: *IJCAI: proceedings of the conference*. Vol. 28. NIH Public Access. 2019, p. 1401.

[123] Z. Han, D. Giger, J. Allspaw, M. S. Lee, H. Admoni, and H. A. Yanco. "Building the foundation of robot explanation generation using behavior trees". In: *ACM Transactions on Human-Robot Interaction (THRI)* 10.3 (2021), pp. 1–31.

[124] S. H. Huang, D. Held, P. Abbeel, and A. D. Dragan. "Enabling Robots to Communicate their Objectives". In: *Autonomous Robots* 43.2 (Feb. 2019).

[125] S. Sreedharan, T. Chakraborti, and S. Kambhampati. "Foundations of explanations as model reconciliation". In: *Artificial Intelligence* 301 (2021), p. 103558.

[126] S. Thellman and T. Ziemke. "The perceptual belief problem: Why explainability is a tough challenge in social robotics". In: *ACM Transactions on Human-Robot Interaction (THRI)* 10.3 (2021), pp. 1–15.

[127] O. Watkins, S. Huang, J. Frost, K. Bhatia, E. Weiner, P. Abbeel, T. Darrell, B. Plummer, K. Saenko, and A. Dragan. "Explaining robot policies". In: *Applied AI Letters* 2.4 (2021), e52.

[128] P. Qian and V. 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. Virtual Event, New Zealand: International Foundation for Autonomous Agents and Multiagent Systems, 2022, pp. 1083–1091.

[129] D. Calvaresi, A. Najjar, A. Omicini, R. Aydogan, R. Carli, G. Ciatto, Y. Mualla, and K. Främling. "Explainable and Transparent AI and Multi-Agent Systems". In: *LECTURE NOTES IN COMPUTER SCIENCE* 14127 (2023), pp. 1–281.

[130] Y. Rong, T. Leemann, T.-T. Nguyen, L. Fiedler, P. Qian, V. Unhelkar, T. Seidel, G. Kasneci, and E. Kasneci. "Towards human-centered explainable ai: A survey of user studies for model explanations". In: *IEEE transactions on pattern analysis and machine intelligence* (2023).

[131] F. Sado, C. K. Loo, W. S. Liew, M. Kerzel, and S. Wermter. "Explainable goal-driven agents and robots-a comprehensive review". In: *ACM Computing Surveys* 55.10 (2023), pp. 1–41.

[132] S. Milani, N. Topin, M. Veloso, and F. Fang. "Explainable reinforcement learning: A survey and comparative review". In: *ACM Computing Surveys* 56.7 (2024), pp. 1–36.

[133] P. Qian and V. V. 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*. 2024, pp. 847–851.

[134] S. Chernova and A. L. Thomaz. "Robot learning from human teachers". In: *Synthesis lectures on artificial intelligence and machine learning* 8.3 (2014), pp. 1–121.

[135] G. Ajaykumar, M. Steele, and C.-M. Huang. "A survey on end-user robot programming". In: *ACM Computing Surveys (CSUR)* 54.8 (2021), pp. 1–36.

[136] J. Zhang, J. Zhang, K. Pertsch, Z. Liu, X. Ren, M. Chang, S.-H. Sun, and J. J. Lim. "Bootstrap Your Own Skills: Learning to Solve New Tasks with Large Language Model Guidance". In: *Conference on Robot Learning*. PMLR. 2023, pp. 302–325.

[137] K. Mahadevan, J. Chien, N. Brown, Z. Xu, C. Parada, F. Xia, A. Zeng, L. Takayama, and D. Sadigh. "Generative expressive robot behaviors using large language models". In: *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 2024, pp. 482–491.

[138] U. B. Karli, J.-T. Chen, V. N. Antony, and C.-M. Huang. "Alchemist: LLM-Aided End-User Development of Robot Applications". In: *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. 2024, pp. 361–370.