Agent-based social skills training systems: the ARTES architecture, interaction characteristics, learning theories and future outlooks

Al Owayyed, M.; Tielman, M.L.; Hartholt, Arno; Specht, M.M.; Brinkman, W.P.

DOI
10.1080/0144929X.2024.2374891

Publication date
2024

Document Version
Final published version

Published in
Behaviour and Information Technology

Behaviour & Information Technology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tbit20

Agent-based social skills training systems: the ARTES architecture, interaction characteristics, learning theories and future outlooks

Mohammed Al Owayyed, Myrthe Tielman, Arno Hartholt, Marcus Specht & Willem-Paul Brinkman

To cite this article: Mohammed Al Owayyed, Myrthe Tielman, Arno Hartholt, Marcus Specht & Willem-Paul Brinkman (20 Jul 2024): Agent-based social skills training systems: the ARTES architecture, interaction characteristics, learning theories and future outlooks, Behaviour & Information Technology, DOI: 10.1080/0144929X.2024.2374891

To link to this article: https://doi.org/10.1080/0144929X.2024.2374891

Published online: 20 Jul 2024.

BEHAVIOUR & INFORMATION TECHNOLOGY
https://doi.org/10.1080/0144929X.2024.2374891

OPEN ACCESS Check for updates

Agent-based social skills training systems: the ARTES architecture, interaction characteristics, learning theories and future outlooks

Mohammed Al Owayyed [a,d], Myrthe Tielman [a], Arno Hartholt [c], Marcus Specht [b] and Willem-Paul Brinkman [a]

[a]Interactive Intelligence, Delft University of Technology, Delft, The Netherlands; [b]Web Information Systems, Delft University of Technology, Delft, The Netherlands; [c]USC Institute for Creative Technologies, Los Angeles, CA, USA; [d]College of Computer Science, King Saud University, Riyadh, Saudi Arabia

ABSTRACT
Agent-based training systems can enhance people's social skills. The effective development of these systems needs a comprehensive architecture that outlines their components and relationships. Such an architecture can pinpoint improvement areas and future outlooks. This paper presents ARTES: a general architecture illustrating how components of agent-based social training systems work together. We studied existing systems and architectures for training and tutoring to design ARTES and identify its essential components and interaction characteristics. ARTES comprises two core components: the agent simulation of social situations, and educational elements to provide guided learning. We link ARTES's crucial components to four primary learning theories (behaviourism, cognitivism, social cognitive theory, and constructivism) to illustrate the role of agent simulation and tutoring elements in establishing desired learning outcomes. Furthermore, we map ARTES's components against eight architectures, 43 systems and three tools to indicate the components' relevance, completeness, generalisation, and deployment potential across contexts. In addition to ARTES, the paper also contributes by identifying future improvements and research directions, such as the agent's thinking, tutoring methods, knowledge transfer, and ethical implications. We believe ARTES can help bridge the gap between virtual human simulations and impactful educational learning, offering training system developers desirable features like understandability and adaptability.

ARTICLE HISTORY
Received 29 November 2023
Accepted 25 June 2024

KEYWORDS
Social skills training; interactive agent; intelligent tutoring system; agent-based training; modular architecture; learning theory

1. Introduction

Social communication skills are important for human interaction. Training these skills can, therefore, be beneficial, including training them with computer applications. Software engineers designing these systems would benefit from a high-level application overview; that is, a conceptual architecture describing the system's components, their interrelationships, and properties. This overview helps engineers identify what is necessary and build applications with desirable characteristics such as understandability for other developers and adaptability (Losavio, Chirinos, and Pérez 2001). Nonetheless, the area of social skills training systems currently lacks such an architecture of the components and their relationships. To this end, we present a conceptual architecture for agent-based social skills training systems. This architecture outlines the general functionalities and components necessary to deliver these functionalities. The architecture is versatile and could be applied to various forms of social skills training, including interviews, negotiations, and taking medical histories. The architecture is also agnostic to the agent's embodiment, modality of interactions, software implementation, and deployment. Because of their importance, it caters explicitly for training on dialogues aimed at changing people's perspectives and beliefs, such as getting vaccinated, becoming physically active, or quitting smoking. Therefore, we specifically focus on simulating the thought process of a communication partner but also on how the architecture facilitates the embedding of instructional principles. Moreover, we also look into new opportunities and further research directions for agent-based social skills training.

1.1. Why social skills training systems

Social skills training is beneficial to train people on problems arising in different social settings, such as social anxiety (Beidel et al. 2014), patient-doctor communication (Oh,

CONTACT Mohammed Al Owayyed m.alowayyed@tudelft.nl Interactive Intelligence, Delft University of Technology, Delft 2628CD, The Netherlands; College of Computer Science, King Saud University, Riyadh 11362, Saudi Arabia

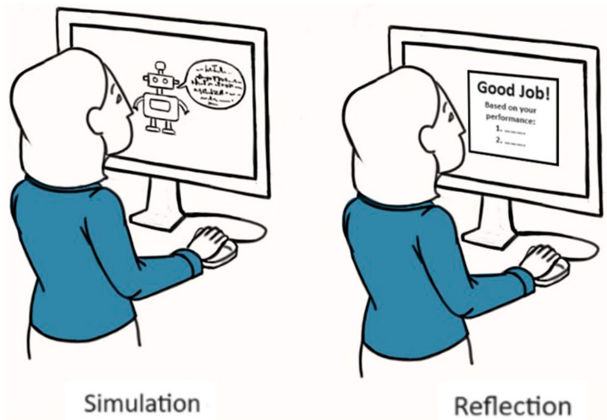


Figure 1. In a typical agent-based social skill training system, the learner (on the left) engages with a virtual agent role-playing as a human, like a virtual patient, to practice a skill. Based on this interaction, the learner receives feedback, as shown on the right.

Jeon, and Koh 2015), and business customer service (Mouawad and Kleiner 1996). Yet, this training is often reported to be costly, time-consuming, and involves several individuals, e.g. training student doctors on breaking bad news (Ochs et al. 2019). A computer-simulated person within a training system offers a practical solution. It role-plays a situation for learners to learn from, whereby they interact with an interactive computer agent representing a human in a specific training scenario (Figure 1). These systems allow learners to practice interactions in a safe, cost-effective, and readily available environment (Gratch, DeVault and Lucas 2016; Mascarenhas et al. 2010; Motgerm, Franch, and Marco 2022).

1.2. Architectures for social skills training systems

Rather than building a training system from scratch, it would be efficient for developers and researchers to leverage an architecture that describes the connection and flow between the different components. Integrating different artificial intelligence (AI) technologies and domain knowledge is crucial to the success of training environments (Kenny et al. 2007), making such a system complex by its nature. Thus, an architecture can help designers abstract and understand such complexity. Several architectures can be adapted for training systems, primarily: (1) tutoring models (e.g. Nwana 1990), which usually focus on the educational process rather than simulations; (2) virtual human architectures (e.g. Lee et al. 2008); and (3) the integration of both (e.g. Core, Lane, and Traum 2014). Other architectures have been proposed in adjacent domains, like in social simulation (Bourgais, Taillandier, and Vercouter 2019). However, these simulations aim to study scenarios involving interactions among many agents that mimic human behaviour, such as simulating the behaviour of many people in an evacuation scenario. This makes them different from social skills training simulations, which focus on simulating social interactions to train human learners.

At its core, training systems are systems to teach (the goal) (J. R. Anderson, Boyle, and Reiser 1985) by having a human interact with a simulated environment (the means). Adding tutoring to such simulations could contribute to a guided learning experience, which is more effective than unguided learning (Clark 2007; Kirschner, Sweller, and Clark 2006; Lee 1999; Titov et al. 2008; Topu and Goktas 2019). Furthermore, computer-based tutoring can be time- and cost-effective compared to in-house tutoring, e.g. debriefing session (Akyuz 2020; Niemiec, Sikorski, and Walberg 1989). However, we noticed that many existing architectures and systems tend to emphasise either the goal, with less utilisation of the simulation potential, or the means, with less emphasis on pedagogical elements. Yet, these elements combine well and could enhance training system benefits in general. Therefore, we argue for an architecture that supports both the goal and the means of training systems.

Furthermore, we need an architecture to describe and develop systems that capture more complex training simulations. The complexity of these systems spans from modelling passive interactions to simulating human cognition. Examples of the more 'complex' systems include negotiation training (e.g. Johnson et al. 2019) and job interviews (e.g. K. Anderson et al. 2013). Other systems in this spectrum aim at training learners to understand and change an individual's cognition or mental state through conversations. These systems allow a human learner to practice how to change a person's beliefs, attitudes, or emotions on a simulated human to equip the learner with skills to apply in real-life interactions with clients. For instance, simulations might focus on de-escalating aggressive scenarios (e.g. Bosse, Gerritsen and de Man 2016) or persuading a virtual patient to seek assistance (e.g. Demasi, Li and Yu 2020). These systems specifically train on altering the simulated human's states, setting them apart from other social skills training. We argue that such systems are among the most intricate to design because developers should replicate components that mimic human thought processes, e.g. modelling changes in human mental states based on a perceived learner's states. An architecture can provide clearer insights into these systems and their components.

Several surveys have explored training systems, including for virtual patients (Battegazzorre, Bottino, and Lamberti 2020) and those utilizing virtual reality (Zahabi and Abdul Razak 2020). The closest to our work is by Bosman, Bosse and Formolo (2018), where they explored 12 agent-based training systems, including those with changing states, and analysed them across six dimensions (e.g. internal reasoning and feedback). Yet, they did not address potential research directions or possible advancements in this area.

1.3. Main contributions

In this article, we present ARTES, i.e. the Agent-based training architecture for social skills. ARTES outlines key components that could ideally be present in a system that simulates a training situation with one agent. Being a conceptual architecture (Clements and Northrop 1996), it describes the components and the connection and flows between them, regardless of implementation decisions. Thus, the architecture provides a valuable reference point for researchers and developers looking to build upon implementations, while not providing technical implementation information as can be expected in an implementation architecture (Johnson and Henderson 2002). One of the key features of ARTES is that it integrates components from two perspectives to maximise learning gains, namely from agent simulations and tutoring systems. In addition, the architecture enables us to identify the important components of training systems, which we can examine further by looking at existing training systems, particularly those focused on changing mental states. By examining the current training systems; offering suggestions for advancements from adjacent fields; linking to learning theories; and discussing future outlooks and research directions, we hope to advance the state-of-the-art in this area.

1.4. Approach

The article follows an approach similar to Harms et al. (2018) and Brabra et al. (2021), where they presented a model and discussed current states and possibilities of applying segments of that model from different dimensions. Similarly, we present an architecture and focus on the essential components of training systems. Although our objective was not to conduct a systematic literature review, we defined the architecture and the characteristics by examining a wide range of literature on agent-based training systems and conversational agents in general. To extract recent examples of such systems, we used different sources: (1) by searching for keywords, such as: 'training system', 'social skills', 'agent', and 'virtual agent', on Google Scholar and Google, (2) by looking at systems mentioned in three literature reviews on social skills training that fits within our scope (Battegazzorre, Bottino, and Lamberti 2020; Bosman, Bosse and Formolo 2018; Ding 2020), (3) by extracting relevant papers published in the Interactive Virtual Agents conference (IVA) between 2019–2022, a period chosen to include also the latest research published after these three literature reviews mentioned, and (4) by looking at reverse citations, where literature references were screened when deemed relevant. The included systems also had to fit within the scope of our architecture, i.e. training individuals in social skills through communicating with an agent. Relevant architectures also inspired ARTES: (a) Architectures of training systems (e.g. Bosse, De Man, and Gerritsen 2014; K. Anderson et al. 2013; Kenny et al. 2007); (b) Virtual agents architectures, mainly for input/output components (Hartholt et al. 2013; McTear, Callejas and Griol 2016), agents' thinking-acting cycles (Kenny et al. 2007) and agents' cognition (Pérez et al. 2016); (c) Intelligent Tutoring Systems (ITS) models (Cheng et al. 2009; Nwana 1990; Siemer and Angelides 1998); and (d) An integration of ITS and virtual humans (Core, Lane, and Traum 2014).

This paper is organised as follows: We first introduce ARTES, an architecture for agent-based training systems for social skills. We then describe the architecture and its components and present the characteristics by which training systems model the interactions between the learner and the interactive agent, a term which we will use interchangeably with virtual agents in this paper. After that, we identify and discuss ARTES's essential components in training systems, followed by showing their role in four learning theories. Finally, we end the paper by mapping ARTES to eight training systems architectures, classifying the examined systems based on the essential components, linking ARTES to development tools, and discussing further research directions in this area. This research contributes by (a) proposing an architecture for agent-based training systems for social skills, which designers and researchers could use to build on their systems; (b) describing current advances in the aforementioned systems and providing suggestions for what could be integrated from other adjacent domains (e.g. conversational agents); and (c) suggesting future research opportunities for agent-based training systems.

2. ARTES architecture and characteristics of social skills training systems

What should an architecture for agent-based social skills training systems look like? What are the components?

And how is interaction with learners characterised? These are the key questions to address now. To explain the abstract concepts in the architecture, we will use the hypothetical example of Lila, a new volunteer in a child helpline. As part of her training, Lila interacts with a simulation system, which includes: (1) a virtual agent simulating a bullied child contacting a children's helpline; and (2) an educational model to manage the learning process. This example is relevant to existing social skills training systems, e.g. to train children in dealing with bullying (Aylett et al. 2005), or to train crisis helpline counsellors (Demasi, Li and Yu 2020). However, it differs in its focus on training a child helpline counsellor to interact with a bullied child.

2.1. Agent-based training architecture for social skills (ARTES)

Figure 2 illustrates the high-level architecture of the virtual training system and how a learner, like Lila, interacts with it. These interactions take place in a communication setting, i.e. a scenario. In Lila's scenario, she interacts with a child who has been bullied and wants the helpline to call their school. This simulated scenario allows Lila to practice applying the helpline's communication protocols, which are the steps by which she can help the child explore viable solutions, e.g. to inform a teacher about the bullying. Therefore, Lila practices facilitating a change in how the child perceives and thinks about the situation. Two essential components are needed to facilitate a change: a simulation of the virtual child, which Lila practices with; and an educational model to guide and teach Lila about communication protocols. For the virtual child simulation, Lila interacts with a user interface that resembles the setting used by child helpline volunteers. This includes a chat interface for exchanging short text messages with the child, access to manuals, and a display showing the number of children waiting in the queue. Behind the interface is the virtual child simulation, represented by a mind and a body. The simulated child's mind aims to sense and understand Lila's input and thinks of a reply based on many factors, such as the conversation's history, Lila's empathy, and its belief in the helpline's ability to help. The virtual child uses its body, the interaction modality, to communicate responses. This could have various representation types, such as a physical robot, a virtual character, or, in the case of the virtual child, generating a response and sending it to Lila via the chat interface. Besides a single modal sense-to-act cycle, multi-modal implementations are also possible, e.g. monitoring Lila's emotions expressed via her facial expressions and text input, and also responding across multi-modalities, e.g. a virtual child showing its emotion via body gestures and text responses.

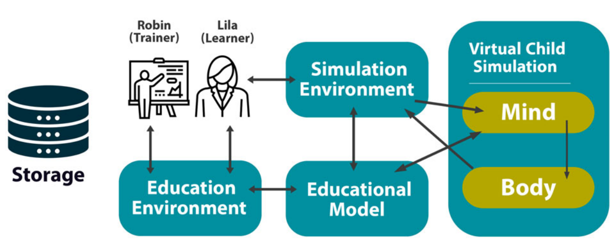


Figure 2. A simplified model of the proposed architecture, using the virtual child example.

During and after Lila's interactions with the virtual child, she might not know what she did right or wrong. To address this issue, the training tool incorporates an educational model that aims to improve Lila's understanding of the child's behaviour, manage the learning process, and guide her in achieving the learning goals. The educational model uses domain knowledge, such as the helpline's communication protocols, Lila's competence level, and the tutor model, to decide what and how to teach Lila. The educational model could assess Lila's progress, adapt the training scenario or child's personality, and provide Lila with guiding feedback. Lila can access the latter through the educational interface. This interface is also used by Robin, Lila's human trainer, to monitor Lila's progress. Robin can also use the configuration interface to influence the training by, e.g. changing the personality of the virtual child.

The proposed full ARTES architecture (Figure 3) aims to present the ideal set of system components that simulate a training situation with one agent, e.g. a virtual child. Lila's example demonstrates that this modular architecture can be tailored to match the simulated training and learning goals. Table 1 lists and explains ARTES's main components.

2.2. Characteristics of learner-agent interactions in training systems

To better understand the interactions between a human learner and an interactive agent in training systems, we identify several key characteristics defining such interactions. Table 2 shows them and aligns them with previous works on taxonomies for multi-modality inputs and outputs (Feine et al. 2019; Turk 2014), dialogue management in conversational agents (Harms et al. 2018; McTear, Callejas, and Griol 2016) and the existing literature reviews on training systems mentioned earlier. These characteristics and their aspects help explain the components of ARTES and understand the differences between various system implementations.

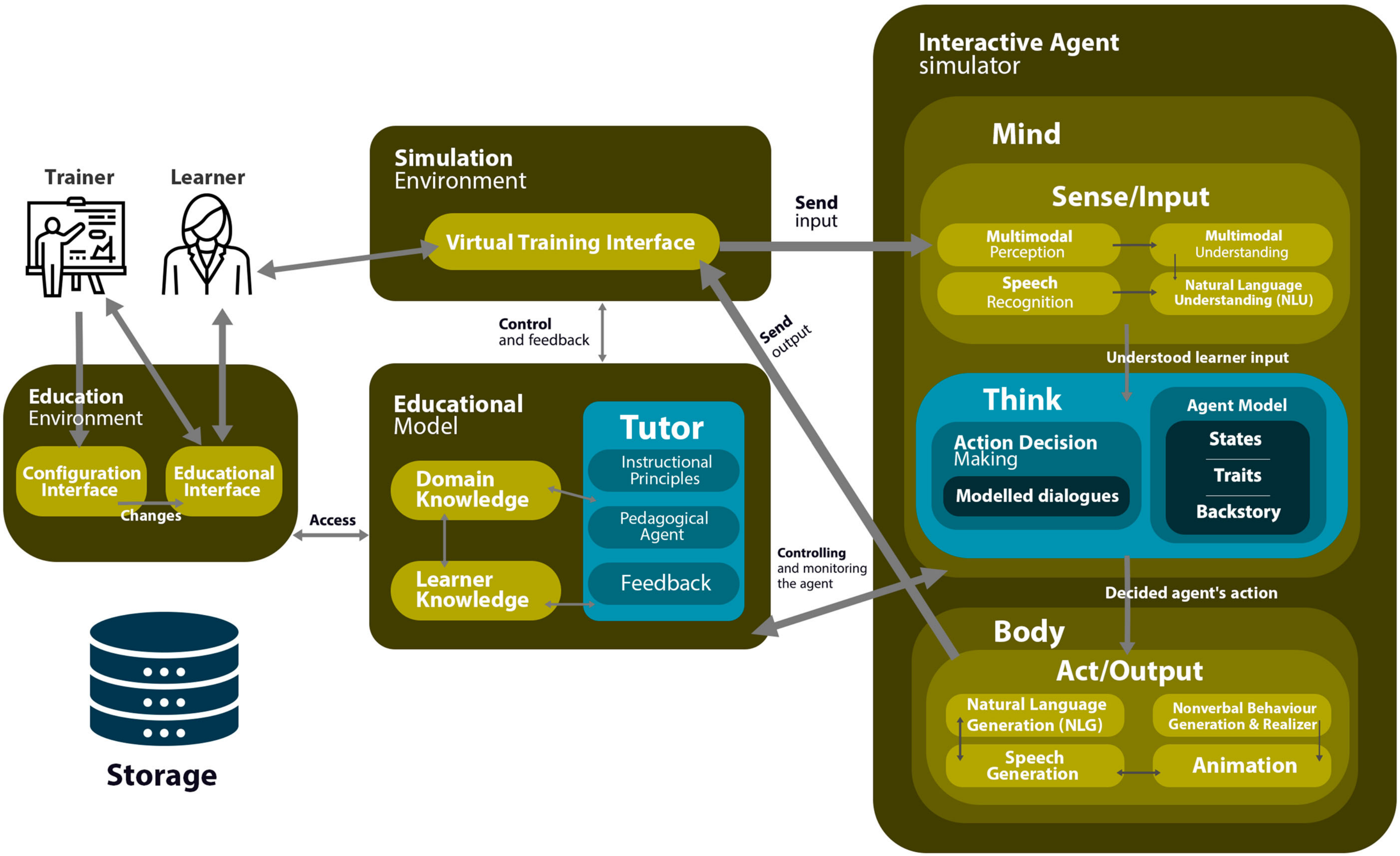


Figure 3. General architecture of agent-based training for social skills (ARTES). The components highlighted in blue, 'Think' and 'Tutor', are crucial for training learners in social skills. The two components are discussed in detail in this article.

Table 1. The main components of the architecture.

Component	Description
Virtual training interface	The simulation environment interface, in which a learner interacts with an agent
Sense/Input	As part of the agent's mind, its main purpose is to sense the user inputs and interprets it. The input could include talking (speech recognition), typing (natural language understanding), happy or sad gestures (multi-modal), or a combination of them
Think	This part of the mind decides on the agent's next action, once the agent has interpreted the learner's input. For this, it might consider modelled dialogues, agent model (e.g. they have a short temper or negative beliefs of the situation), and how the learner is progressing in training (e.g. the agent takes a less cooperative stance to challenge more experienced learners)
Act/Output	Part of the agent's body. After the agent decides on an action, it 'acts' by, e.g. showing sadness (animation) or raising voice (speech generation), which is shown in the training interface
Domain knowledge	Contains information concerning the training subject and learning goals. It includes facts, procedures and rules of the topic, such as how to follow a certain communication protocol
Learner knowledge and skill	The learner's current knowledge and skill state in the context of the full domain knowledge-skill set, e.g. to what extent the learner is able to apply a certain protocol
Tutor	The tutor manages the system's teaching process by considering domain knowledge and the learner knowledge and skill components. The tutor decides on what knowledge or skill to focus on, how to do this, and provides feedback on the learner's performance.
Educational interface	The learner and trainer interact with the educational interface to check the learner's progress and interaction
Configuration interface	This interface allows a trainer to control the simulation or the educational model, e.g. by changing the agent's mind configuration or tasks
Storage	The storage keeps data related to the system, and its uses, such as the chatlog or learner's performance data

While any combination of these characteristics can theoretically occur, certain patterns are more likely based on factors such as the agent's roles, technology used, and training scope. When considering the case of Lila's interaction with the virtual child, the interaction takes place through open written inputs, where Lila writes in a conversation text field. The child's response is provided through verbal textual output on the interface, and the reasoning is based on a hybrid approach between rule-based and data-driven, allowing for training in a safe, explainable, and controlled environment. The child also has defined internal states and beliefs that change while talking with Lila, as well as a background story of being bullied at school, and memory of past conversations. The trainer, Robin, can modify the virtual child's configuration by changing its belief values, making the child more or less adamant. In other cases, the agent simulation may have different

Table 2. Characteristics of the interactions between a learner and a simulated agent.

Characteristic	Aspect	Definition
Input mode	Verbal	Written learner input
	Visual	Any visual input such as facial expressions and gaze
	Sound	Sound input (e.g. speech)
	Haptic	E.g. pressure, selection
	Others	Sensors, emotion recognition, motion capture
Verbal interactions input	Closed	Options for the learner to choose from as a reply
	Open	Free text speech input
	Both	Have both types of input
Reasoning	Rule-based	The rules of the virtual agent interactions are predefined
	Data-driven	The systems learn the rules from conversations (e.g. a dataset)
	Hybrid	Combination of rule-based and data-driven
Internal state	Defined model	The agent has defined states on how they affect the interactions
	Limited	The agent has one or two parameters (e.g. trust level) that change
	None	No explicit representation of internal states
Agent knowledge	No history	No memory of past experience is registered
	Background only	The agent has a history of its backstory
	Dialogue only	The agent has a memory of past interactions
	Dialogue and background	The agent has a memory of past interactions and their background
Trainer involvement	Not involved	No human is involved and the agent interactions are automated
	Configuration	The trainer monitors interaction and configures educational settings
	Human-initiative	Automated interactions, but the trainer can intervene and control
	Agent-initiative	The agent asks the trainer to take control (e.g. making a decision)
	Wizard of Oz	A human understands the input and controls the output of the agent
Output mode	Verbal	The agent utterances
	Visual	Any visual output appearing on the agent (e.g. animations)
	Sound	Sounds the agent outputs (speech or non-speech)
	Others	Outputs that can not be seen or heard (e.g. haptics)

characteristics. For example, a visual embodiment of a virtual recruiter for training on non-verbal cues, direct rule-based reasoning for answering questions in history-taking, or an agent-initiated involvement to demonstrate specific cues of a virtual patient with depression.

3. Key social skills training components

The ARTES architecture includes components that are also included in non-training systems, such as multi-modality or dialogue systems. As these more generic components have been explained elsewhere (e.g. Harms et al. 2018; Turk 2014), we will now focus our attention on the 'Think' and 'Tutor' components. Both components play a pivotal role in training learners' social skills to understand and change an individual's emotions, cognitions, and behaviours consequently. Table 3 shows the sub-components of these two components.

3.1. Think components

Most training systems we studied were limited in their simulation of the agent's thinking. We believe there is a potential for improvement, as by showing the change in the agent's belief states, learners can construct a mental model of the agent's thinking process, from which they can learn the consequences of their actions. In other words, they can construct a mental model of the situation they operate in (Jonassen and Henning 1999). For Lila, this means she could understand how the child's emotion and thinking changes when she follows the helpline protocol. In this section, we will discuss the elements that affect the agent's action decision, reflecting on the proposed architecture.

3.1.1. Sources of input

Looking back at the model in Figure 3, we distinguished three information sources upon which an agent bases its actions: what Lila has said (information provided by sense/input component); what the child is thinking about itself or the world (the agent model); and what is an appropriate learning experience for Lila (input from the educational model). Let us start with the first one. The agent considers Lila's input based on what the two have said so far, e.g. did the child pose a question, and is this Lila's answer? If the child is happy and trusts Lila, the agent might consider her answer instead of rejecting it. In other words, the agent's decision depends on the agent's cognitive and emotional states, i.e. the second source of input. So, if Lila showed empathy, the child's trust in Lila would increase; thus, the agent would be more willing to open up. Finally, Lila might be cognitively and emotionally ready for a more challenging twist in the scenario. Sensors indicate she is relaxed, and the learner knowledge and skill information indicate she has sufficient understanding for the next step in her training. Therefore, the educational model might change the parameter setting of the Action Decision Making component on the fly, for example, making the child more susceptible to any suggestion of rejection by Lila. This third source, therefore, influences the agent's decision-making on a meta-level, by changing the rules it applies.

3.1.2. Basic decision making

With the three input sources, an agent's thinking could vary widely, from being simple options-based, matching inputs to direct outputs, to having a complicated structure of decision-making that includes dialogues, goals, and emotions that adapt to the learner's state. A direct thinking process might be suitable for training systems where a change in an agent's beliefs or emotions as part of the conversation is less relevant to the agent's decision-making. For example, to train student doctors in asking specific appropriate questions and collecting information for diagnosing (Shah et al. 2012). However, such dialogues are experienced as being static and inflexible (Bosman, Bosse and Formolo 2018). Therefore, some proposals have been made to tackle this. For example, Hirumi, Kleinsmith, et al. (2016) structured the questions that

Table 3. Overview of Think and Tutor sub-components, which we examine further in this article.

Examined components	Description
Think:	An agent decides on an action based on the understood input
a) Action decision making	Controls how the training scenario proceeds, i.e. how an agent decides what their next step would be based on the overall interaction structure
b) Agent model	Describes the model of the agent with different parameters and states. This could include its emotions and cognition about itself, the learner, and the world. This, in turn, influence the agent's future behaviour
Tutor:	What educational information to present and how
a) Instructional principles	A training system includes principles to teach knowledge, which could affect the agent's interactions with the system. It is related to learning techniques and the learner's state
b) Feedback	Generates and displays feedback to reflect on the learner's performance, e.g. being empathic, or to explain the agent's behaviour, e.g. the changes in agent beliefs
c) Pedagogical agent	A virtual trainer agent guiding or reflecting upon the learner's interaction with the simulated agent. It can have basic or complex interactions, e.g. with the 'thinking' elements

could be asked by learners into topics, e.g. about the patient's medical or social history. Dialogue managers (Harms et al. 2018) go beyond option-based solutions, as they can consider the dialogue history and state when deciding on an action. Take, for example, the following interaction: '[Child] I'm bullied at school, can I talk about that?; [Lila] Sure, I am here to listen; [Child] Thanks. Do children often call about bullying?; [Lila] yes, you are not the only one'. Here, the dialogue manager is running a script and, based on Lila's input, gives an appropriate response and moves on. Another extension is to add one or two state variables about agents, e.g. a trust meter (Schoenthaler et al. 2017). It is an initial attempt to have learners consider the consequences of their actions.

3.1.3. Cognitive models

More complex agent models can be used to elaborate the thinking, e.g. by simulating a virtual child's mental model that reaches decisions based on interactions with Lila. These models are often referred to as cognitive models that simulate human decision-making processes (Taatgen and Anderson 2010). These cognitive models can be application-specific or can be built upon general theories. Application-specific cognitive models, such as a virtual suspect for interrogations that bases its answers on interpersonal relations (Bruijnes et al. 2015), have the advantage of addressing unique simulation aspects, e.g. states which might not be included in general theories. However, they require more work from domain experts to simulate thinking, agent states, and their interconnections. In the case of the virtual child, this involves identifying relevant thinking processes for training, potential cognitive and emotional states for the child, and how these two can be linked together. On the other hand, general cognitive models describe how the agent's state varies, allowing developers to build upon a predefined flow. One well-known and utilized cognitive theory in conversational agents is the belief-desire–intention (BDI) model. This model is built on three aspects: (1) the agent's beliefs about themselves, the learner (i.e. Theory of Mind Leslie 1994), and the world; (2) the agent's desires regarding their goals; and (3) their intentions as the actions they wanted to execute (Georgeff et al. 1998). For instance, when Lila shows empathy towards the virtual child, the child's belief in Lila's understanding of their situation increases, making the child more open. Employing BDI in such a model could simulate human decision-making, adapt to changes in the environment, and directly explain the child's decisions and changes in internal states. In their survey of BDI agents for social simulation, Adam and Gaudou (2016) argued that BDI models are well-suited for agents in training systems. However, they also discussed drawbacks, with the most relevant for training systems being the need for expert input to model agents' values. BDI has been utilized in several training agents (e.g. McShane et al. 2009; Muller et al. 2012; Oijen, Doesburg, and Dignum 2010), and proposed for others (e.g. Baptista et al. 2014; Van der Zwaan, Dignum, and Jonker 2012). The model's flexibility allows for extensions to include social concepts in the agent's mind, such as agent trust (Adam and Gaudou 2016), which would be valuable to train learners on specific social skills and how such concepts affect the agent's decision-making.

3.1.4. Cognitive architectures

Cognitive architectures offer another approach to modelling human cognition in training settings. Although some cognitive architectures are based on BDI, researchers differentiated BDI from cognitive architectures in that the former is based on a philosophical take on behaviour, while the latter considers describing lower-level cognitive processes to simulate a human mind (Adam and Gaudou 2016). Therefore, we will consider them both as methods of cognition modelling in the agent's thinking. Cognitive architectures aim to achieve general human intelligence capable of reasoning, insight, behaviour adjustment, and self-reflection (Kotseruba and Tsotsos 2020). Architectures vary in decisions about actions, typically based on factors from their internal memory and interactions, such as physiological needs, utility, state, and relevance. However, we found limited social skills training applications based on surveyed cognitive architectures (Kotseruba and Tsotsos 2020; Ye, Wang, and Wang 2018). One exception is a 2003 implementation of Soar (Laird 2019) to train army personnel to make decisions in a critical social situation (Hill et al. 2003). The limited examples might be because cognitive architectures aim to simulate general human intelligence and not to 'be an actor' with emotions and states in a specific role-play training scenario. Moreover, they are not as easy to implement as BDI (Adam and Gaudou 2016). While cognitive architectures may not be the primary choice for training systems with changing agent states, they might have a potential added value when continuous adaptability or self-reflection is needed, depending on the agent's past experiences. For example, when building a general virtual human for helpline counsellor training that can be utilised in children's helplines.

3.1.5. Modelling emotions

So far, we have only discussed the cognitive part of the agent's mind. An additional layer to the agent model in a training system is modelling the agent's emotions, often referred to as affective architectures (Pérez et al. 2016). These architectures extend cognitive models, such as BDI, to include more elaborated states or traits to the agent, e.g. by adding emotions, personality, attitude, and mood (Ojha, Vitale, and Williams 2021). As a result, the agent's emotions can influence their decisions and beliefs, leading to varied reactions based on their feelings. Incorporating such feelings in the virtual child example can help Lila understand the child's internal states and emotions and how the interactions affect them, e.g. when the child might feel frustrated and how to address it. Emotions also enrich the simulation, particularly when the learning outcome involves understanding or changing the agent's emotions. One example is TARDIS (K. Anderson et al. 2013), which has an affective architecture. In this job interview training system, the virtual recruiter's emotions impact its decision-making, e.g. becoming more aggressive as they feel angry. Numerous affective models, primarily based on BDI (Ojha, Vitale, and Williams 2021), integrate cognitive modelling with emotions to build a training agent. For instance, EMA (Marsella and Gratch 2009) was employed in a stressful decision-making scenario (Swartout et al. 2006), and FAtiMA (Dias, Mascarenhas, and Paiva 2014) to build a job interview scenario (Mascarenhas et al. 2022).

3.1.6. Data-driven approaches

Agents in training systems usually have rule-based reasoning. Alternatively, data-driven approaches, i.e. a model trained on corpora of conversations (Harms et al. 2018), may also prove helpful. There are many challenges, though, to having data-driven training systems, including the scarcity of conversational data tailored to the simulated scenario, concerns about data accuracy, and technical limitations such as limited understandability of human emotions. For example, building a data-driven virtual child would raise questions regarding sensitivity, availability, and privacy when collecting conversational data from children and helpline volunteers. Human-in-the-loop machine learning could offer a solution when a human trainer acts as the agent and responds to learners' interactions to create a corpus. In Lila's case, this would involve Robin controlling the virtual child. This method, akin to what was used in a storytelling chatbot (Jackson and Latham 2022), could help train the virtual child on actual data for different background stories and interactions. However, it requires Robin's commitment and manual effort to define the child's states in order to develop a human interpretable agent model. Reinforcement learning offers another alternative, where the model of the agent is trained based on user interactions, and the agent's correct behaviour is rewarded (Williams 2008). While reinforcement learning has been utilized in related domains, such as healthcare assistants (Yasavur, Lisetti, and Rishe 2014), implementing this approach in a training system like Lila's may pose challenges. These challenges can arise when an agent needs to have consistent and predictable agent behaviour across different learners. Furthermore, defining suitable reward functions for simulation in such a sensitive scenario, which requires interdisciplinary expertise, poses its own hurdles. Therefore, human experts could check such rewards to determine if the virtual child exhibits appropriate behaviour during the communication protocol phases or to assess whether Lila gained the correct learning skill, such as showing empathy to the child.

3.1.7. Trainer's involvement

Human trainers could also influence the agent's decision-making process. ARTES captures this by incorporating the interaction flow from the trainer, in our case, Robin, to the educational model and then to the agent's thinking. Aside from the Wizard of Oz approach, where a human trainer substitutes the entire agent cognition, Table 2 illustrates two ways in which a human trainer can impact the agent's action decision: based on human or agent initiative. Human initiative occurs when the trainer acts as a supervisor guiding the agent and intervening when necessary, for example, if the virtual child does not exhibit the expected behaviour or if Robin wishes to control parts of the agent (e.g. Hartanto et al. 2014). On the other hand, agent initiative has the agent asking Robin to choose on its behalf, substituting a part of the agent's cognition with the trainer's input. The latter is used when the agent is stuck in a conversation or faces a moral dilemma (van der Waa et al. 2020). While this approach requires a human trainer on standby during the training session, it can also be utilized in a data-driven approach to training the agent model, where the agent learns from the trainer's decisions by requesting their input.

3.2. Tutor components

The tutor model is an intermediate between domain knowledge and learner knowledge in the educational model. The tutor manages the learning experience by supervising learning tasks, providing feedback, and maintaining learner engagement and motivation (Bourdeau and Grandbastien 2010). Unlike agents with their human-like manifestation, tutors in intelligent tutoring systems could manifest themselves in many forms. These can be colour-coded feedback indicators, such as red and green symbols showing the learner's performance or topic recommendations. Many models have been reported for tutors, detailing their logic and presentation style for giving feedback or guidance to a learner (Alkhatlan and Kalita 2018). Having such a model in a training system is helpful to educate learners, as the lack of guidance could leave the learner only with the simulation experience, limiting its effectiveness (Hattie and Timperley 2007). Consider, for instance, if Lila directly instructs the child to ask the school for help. Such an approach contradicts the helpline's guidelines (de Beyn 2003), where a counsellor is expected to coach the child to come up with that solution rather than telling the child directly what to do. A tutor could intervene in such cases, showing Lila how her approach deviated from the helpline's guidelines.

Adding a tutoring model could enhance the training impact by giving the right feedback at the right moment, thereby integrating guidance naturally into the training system's teaching approach. In short, this model is essential for offering a guided learning experience. Designers could look at a number of learning theories (Ertmer and Newby 1993) established by different schools of thought to utilise them in training systems, thus, facilitating the learning process. On the other hand, training systems are typically designed to teach a specific set of skills. Therefore, linking the learning theories and training systems components clarifies the latter's roles, allowing for a more targeted implementation. But before exploring this link, we must first understand the tutor's subcomponents.

3.2.1. The three subcomponents

The tutor model (Figure 3) is responsible for teaching the learner, and has three components: instructional principles, a pedagogical agent, and feedback. The first component, instructional principles, or as some refer to it, instructional strategies (Weston and Cranton 1986), entails psychological techniques the training system can employ to deliver the training. Take, for example, the scaffolding strategy (Van Der Stuyf 2002). Here, the system initially offers Lila extensive help, then gradually reduces support as she gains experience. Next is the pedagogical agent, who behaves as a social actor as it simulates a teacher guiding the learner, e.g. as a person offering guidance and assistance to Lila. Although pedagogical agents are perceived positively in training systems (e.g. Murali et al. 2021, 2022; O'Brien et al. 2019), they are often not included. For example, Table 6 shows that in 43 systems we studied, only 12 had a pedagogical agent. The third tutoring component is feedback, which can give Lila insight into her performance. The feedback can be based on the modelled knowledge, e.g. to tell Lila what she did wrong and offer hints, or the agent's states, e.g. why the virtual child does not trust Lila. Yet, most examined systems primarily focus on the former (Table 6).

The tutor model can interact with (a) the agent simulation, e.g. for monitoring belief changes in the simulated child that needs explaining, (b) the domain and the learner knowledge, e.g. for determining when Lila's training should move on to a more challenging situation, and (c) the human teacher, e.g. Robin configuring the tutor. Similar to ITSs (VanLehn 2006), the tutoring model operates within two learning cycles: the within-session cycle, e.g. for providing hints during a session, and the between-session cycle, e.g. for deciding on new scenarios.

3.2.2. Linking to learning theories

Tutor components can represent and facilitate the implementation of learning theories, thus supporting their incorporation and abstraction within ARTES. Four prominent learning theories exist today that explain how skills are acquired: behaviourism, cognitivism, social cognitive theory, and constructivism (Ertmer and Newby 1993; Schunk 2012). Table 4 highlights the roles of four software components within the context of these learning theories. They are the three tutor components and the virtual agent simulation. The first software component, instructional principles, derives its role directly from the core ideas of learning theory, subsequently guiding the design of agent simulation. Take, for example, behaviourism. The instructional principles focus on forming appropriate stimuli-response relations; thus, the simulation's role is to offer realistic stimuli. This flow similarly applies to the roles of the pedagogical agent and feedback, which, in our example, work to reinforce desirable responses to stimuli through guidance and feedback. Now, let us look at the learning theories in more detail.

Behaviourism can be applied to teach facts and tasks through repetition and memorization. This approach supports reinforcement and guidance in accordance with instructional principles. Therefore, the realism of the agent's actions and consequences are important as they establish stimuli for which a response needs to be trained. To strengthen this stimulus-response, the pedagogical agent and feedback components should reinforce positive or negative behaviours. Some examples of adopting behaviourism are showing angry responses and allowing scenario repetition following incorrect actions by the learner (Guetterman et al. 2019), as well as a virtual coach giving a thumbs-up for correct actions (Murali et al. 2021).

Table 4. The roles of ARTES components within four primary learning theories. These components comprise the three tutor elements (instructional principles, pedagogical agent, and feedback) and the agent simulation. The roles explain how these components fit within each learning theory.

Learning theory	Instructional principles	Agent simulation	Pedagogical agent	Feedback
Behaviourism: Reinforcement to shape desirable behaviours	Use of techniques to guide learners behaviour	Realism of stimulus	An entity that provides feedback or assistance	Provide reinforcement feedback
Cognitivism: Focus on mental process to acquire knowledge	Use of techniques to guide learners knowledge development	Simulating several situations to learn from	Teaching and explaining	Explaining situations and learner's decisions
Social cognitive theory: Learning through social observations	Offer suitable examples of interactions	Simulating several situations to learn from	Interacting with the agent simulation	Explaining situations
Constructivism: A learner constructs their own understanding and knowledge	Facilitate the learning environment	Explorative situation	Mentoring	Used alongside instructional principles to guide learners

Cognitivism focuses on constructing a learner's cognitive process and knowledge, which requires an understanding of the relevance and application of targeted skills. Implementing instructional principles, such as information chunking, can build this understanding. Accordingly, an agent simulation with several learning situations is needed to demonstrate the use of a targeted skill, which is further explained by a pedagogical agent. Feedback can then be provided based on the learner's interactions with the simulation. Examples of cognitivism in training systems include: adapting a scenario difficulty to the learner's current knowledge (Bosse, Gerritsen and de Man 2016), a pedagogical agent that explains the next task (Murali et al. 2022), or receiving reflection about the learner's performance (K. Anderson et al. 2013).

The social cognitive theory emphasises learning through observation of peers or coaches. In this perspective, the instructional principles could demonstrate successful interaction with a virtual agent to train the learner. Therefore, the examples should cover different learning situations. Several systems have implemented this approach, including a pedagogical agent that provides examples of taught skills (Murali et al. 2022), or a system that allows learners to view video recordings of mastered skills (Tanaka et al. 2016).

As for *constructivism*, learners should be able to construct their knowledge, e.g. through self-reflection (Korsgaard 1996) or discovery learning (Mayer 2004). In training systems, this translates to exploring virtual agents' scenarios to create their interpretations and communication styles and learning from mistakes. This approach requires a diverse and rich set of scenarios in the simulation. At the same time, the instructional principles can encourage exploration by facilitating and personalising the environment presented to learners. Examples of applying this theory include providing a 'practice' mode to support scenario exploration and learning (K. Anderson et al. 2013), a virtual coach providing insights into the virtual patient's thinking (O'Brien et al. 2019), and offering guiding feedback by highlighting important discoveries during the interaction with the agent (Kleinsmith et al. 2015).

3.2.3. Alternative tutor structure through data-driven approaches

Utilising data-driven techniques in the educational model makes it harder to distinguish between the different educational model components. One example is the Korbit learning platform (Kochmar et al. 2022; St-Hilaire et al. 2022). Korbit aims to provide instruction in various subjects and skills, employing a conversational interface that utilizes reinforcement learning and machine learning to personalize the AI tutor's pedagogical intervention. It utilises machine learning to generate hints based on learner questions. It also retrieves explanations from sources like Wikipedia. However, this technique requires extensive data sources that are typically only available for general educational topics, making it less suitable for training systems that focus on specific domain aspects and train learners accordingly. Nonetheless, Korbit's approach can be valuable, allowing for adapting instructional principles based on learner profiles and previous interactions with the system. This requires a more elaborate training system with diverse scenarios or tasks to build the learner's profile and history. Among the examined training systems, only one system integrated reinforcement learning into its tutoring approach (Georgila et al. 2019), where it manages feedback and prompts self-reflection on incorrect user inputs. This implementation showcases a merging between different components from our model, specifically the instructional principles, feedback, pedagogical agent, and learner knowledge.

4. Relation with the examined systems, other architectures and tools

Having described ARTES, it is time to see how well the architecture captures and resonates with existing

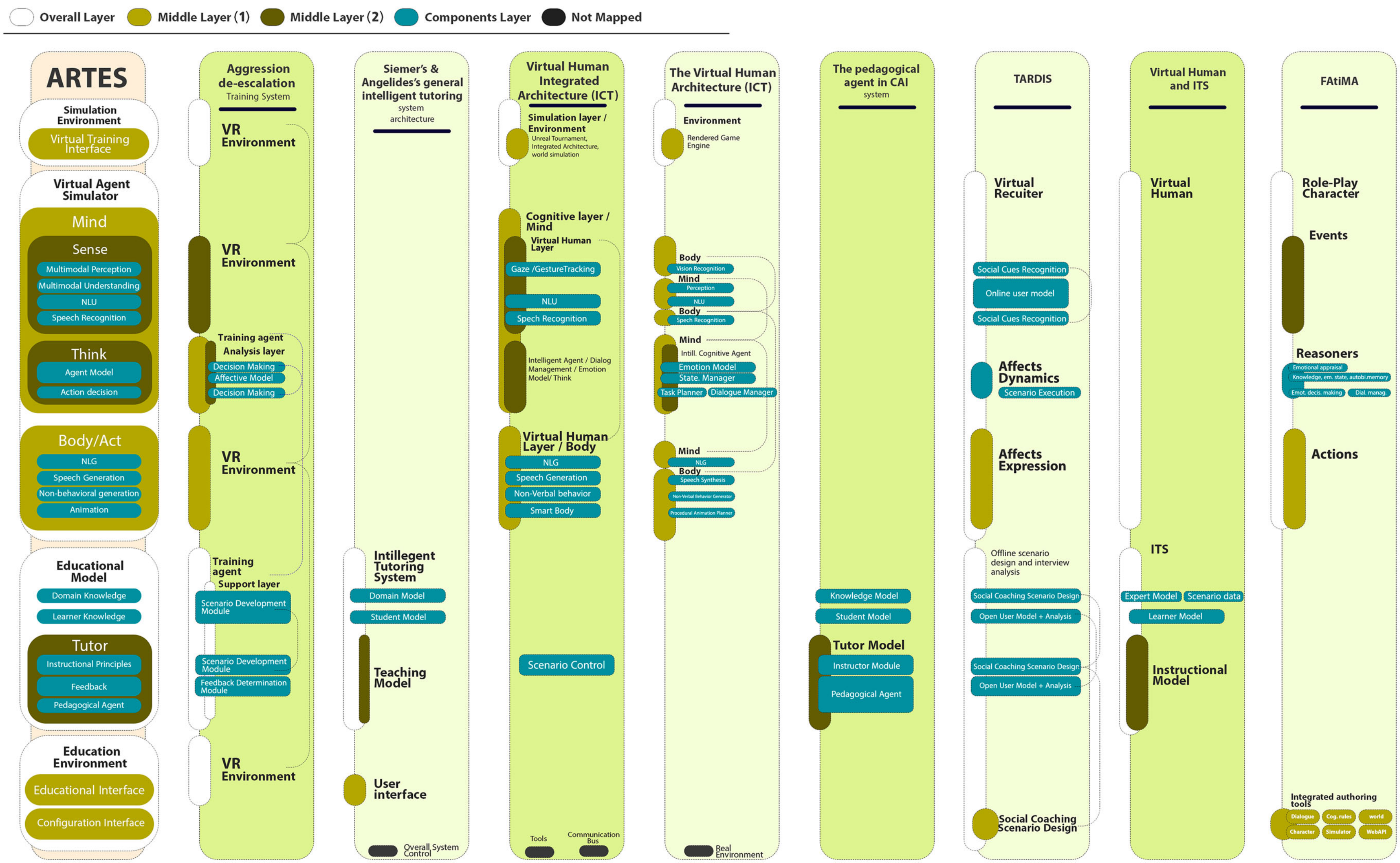


Figure 4. Mapping components between ARTES and eight other architectures. Columns represent different architectures, while rows show ARTES components (on the left) and their counterparts in other architectures. Dotted lines connect component names that map to multiple ARTES components. Colours indicate layers and black areas highlight components absent in ARTES.

systems and architectures, supporting that ARTES is versatile and agnostic. This assessment provides insights into its deployment and adaptability across different situations. Specifically, whether ARTES shows relevance and generalisation and whether its components match those of other architectures. Moreover, looking at how the components were implemented, e.g. if the input type was choice-based or open-ended, gives an insight into how concepts discussed earlier are actually implemented in systems. Next, we will examine how well current software tools support ARTES-based system development. All these points are, therefore, systematically examined here.

4.1. Mapping ARTES to social skills training architectures

How complete is ARTES in comparison to existing architectures? To answer this question, we evaluated the coverage of ARTES against eight architectures suitable for training systems. We did this by mapping the components of the eight architectures to those of ARTES (Figure 4). The mapping offers insights into components commonly defined in training or tutoring systems, compares ARTES's components with other architectures, and highlights any component that ARTES might have overlooked. The eight architectures we analysed are all conceptual architectures deemed relevant to ARTES and applicable within training systems. These architectures are (a) two implementations of training systems [aggression de-escalation (Bosse, Gerritsen and de Man 2016), and TARDIS (K. Anderson et al. 2013)], (b) three virtual agent simulations [virtual human integrated architecture (Kenny et al. 2007), the virtual human architecture (Lee et al. 2008), and FAtiMA Toolkit architecture (Mascarenhas et al. 2022)], (c) two tutoring systems [the general ITS architecture (Siemer and Angelides 1998), and a pedagogical agent in an ITS system (Cheng et al. 2009)], and (d) an architecture integrating ITS with virtual agents (Core, Lane, and Traum 2014). Although the first two architectures were developed to describe specific systems, we think their information is general enough to be applied to training systems. Still, the mapping accuracy is limited by the clarity to which the architectures were described in the literature and potential individual interpretation bias. To mitigate the latter, two coders with a computer science background mapped the eight architectures to ARTES independently, which resulted in a substantial agreement (Cohen's $\kappa = 0.7$) (Landis and Koch 1977). The coders then discussed disagreements to reach a consensus.

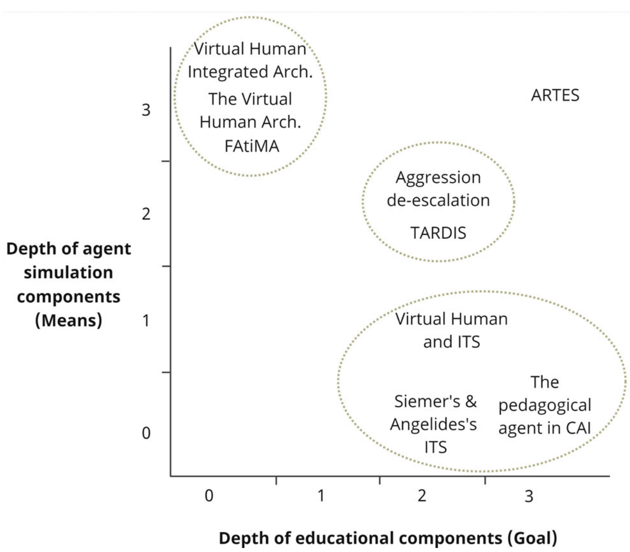


Figure 5. Assessing the level of depth for the eight mapped architectures (Figure 4). We determined the level of depth by calculating the median depth level of the components of an architecture for the two dimensions: the agent simulation model and the educational model.

We could directly map all ARTES components and layers to at least one component from the eight architectures, indicating that ARTES components are relevant to training systems. This gives more credibility to ARTES's components. On the other hand, we did encounter six components from the mapped architectures that did not have suitable matches in ARTES. Two reasons for this are: (i) the components were describing the interaction process in a concealed 'black box' that generalised across distinctive components and their relations [*system control* that handles interactions between all components (Siemer and Angelides 1998) and *communication bus* between all components (Kenny et al. 2007)]; and (ii) that they were from other architectural perspectives than conceptual architecture [a component for the general software *tools* (Kenny et al. 2007) and *real-world environment* that represents the real world (Lee et al. 2008)].

On average, ARTES matches or exceeds the detail level of the mapped architectures, such as FAtiMA, which has more configuration components but fewer Mind and Body components. This also includes specialised architectures like Virtual Humans for agent simulation and ITS architectures for educational systems. Figure 5 shows the depth these eight architectures cover, based on the mapping in Figure 4. There are three main clusters in the figure: (1) in the top left corner are architectures that mainly focus on the agent's simulation (the means); (2) in the bottom right corner are those focusing on the educational model (the goal); and (3) in the middle, the integration of both in well-defined systems. However, none of these eight

Table 5. Categorisations of the different implementations for the examined components, based on the discussion in Section 3.

Components				
Action decision making	Agent model	Feedback	Instructional principles	Pedagogical agent
Decided by: (1) The teacher actor (2) Direct input to output mapping. See Figure A1(a) (3) A scenario sequence is defined. See Figure A1(b)	**Structure:** (1) Fixed changes in the states (2) Limited states change. See Figure A1(c) (3) Representation of changes in many states. See Figure A1(d)	**Moments of feedback:** (1) During interactions (2) After interactions	**Learning theories:** (1) Elements of behaviourism (2) Elements of cognitivism (3) Elements of social cognitive theory (4) Elements of constructivism	**Representation of a trainer:** (1) Through text or sound (2) Embodiment
Also affected by: (1) Agent model (2) Educational model	**Represented states:** (1) Changing cognition (2) Changing emotions	**Content about:** (1) The agent's mind (2) The modelled knowledge	**Extent:** (1) Revolves around one dimension of principles (2) Multiple dimensions	**Pedagogy format:** (1) Static (2) Interactive agent
Input type: (1) Choice-based (2) Open-ended	**Model of:** (1) The agent itself (2) The learner (3) The world	**Covers:** (1) One session or task (2) Multiple sessions or tasks		**Role:** (1) Reflecting and feedback (2) Learning, e.g. describing concepts or answering questions (3) Assisting (4) Social actor

architectures addresses both aspects as deeply as ARTES (located in the top right corner), indicating that these architectures often focus more on either one of the dimensions.[1] In practice, training systems like TARDIS and aggression de-escalation often incorporate components from both educational and agent simulation models, supporting our vision to merge tutoring and agent models. Thus, ARTES demonstrates coverage with both specialized and broader system architectures, underscoring its completeness.

4.2. Examined systems

To further explore the generalisability and relevance of ARTES components, we examined 43 training systems. Specifically, we studied if we could identify the five ARTES important components, as highlighted in Section 3, within the examined training systems, and if the concepts presented in this article resonate with existing systems. To do that, we formulated Table 5, which presents the categorisation criteria for components' properties, such as input type, learning theories, and pedagogy format. We based these categorisations on the ideas we discussed previously in Section 3. Also, the Appendix visually explains in more detail two classifications of the 'decided by' property of the Action Decision Making component (Figure A1(a,b)) and the 'structure' property of the Agent model component (Figure A1(c,d)). Then, we applied the categorisations Table 5 to the examined systems, which resulted in Table 6. Just as with mapping the architecture components, we cannot guarantee the accuracy of our interpretations based solely on the descriptions of these systems. Therefore, we conducted reliability testing to examine this bias. Five secondary coders classified a total of 41 of the 43 systems, with the remaining two systems used to train them in coding. There was moderate inter-rater reliability between the first coder – who assessed all systems – and the secondary coders (average Cohen's $\kappa = 0.55$, range: 0.49–0.63) (Landis and Koch 1977). The coders discussed the ratings to reach a consensus.

Table 6 maps the five components across the examined systems. For Action Decision Making, 49% of systems utilized a scenario sequence to guide agent actions, while 42% employed input-output processing. Additionally, 28% of agents incorporated a representation of internal states. Surprisingly, only 70% of the systems integrated feedback components, where 14% provided feedback only during the interaction, 33% only post-interaction, and 23% during both phases. Regarding instructional principles based on learning theories, Behaviourism was most prevalent at 81%, followed

Table 6. Categorising the examined training systems based on Table 5.

System	Training type	Examined components: Action decision making	Agent model	Feedback	Instructional principles	Pedagogical agent
Bosse, Gerritsen and de Man (2016), Bosse, De Man, and Gerritsen (2014)	Aggression de-escalation	3, 1,2, 1	3, 1,2, 1,2,3	2, 2, 1	1,2,4, 2	–
MPathic-VR (Guetterman et al. 2019; Kron et al. 2017)	Medical communication	3, 1	1	2, 2, 1	1,2, 1	–
Murali et al. (2022)	Counseling	3, 1	1	–	1,2,3 1	2, 2, 2
TARDIS (K. Anderson et al. 2013)	Job interviews	3, 1,2, 2	3, 1,2, 1,2,3	2, 2, 1	1,2,4 2	–
NERVE (Hirumi, Johnson, et al. 2016; Hirumi, Kleinsmith, et al. 2016)	Interview and diagnosis	2, 1,2	1	1,2, 2, 1	2,4, 2	–
ASST (Tanaka et al. 2016)	Communication skills	2, 2	1	2, 2, 1	2,3,4, 2	–
CuCoMaG (Doberstein et al. 2016; Othlinghaus-Wulhorst, Mainz and Hoppe 2019)	Customer service	3, 2	1	2, 2, 1	1,2, 2	–
Virtual suspect William (Bruijnes 2016; Bruijnes et al. 2015)	Suspect interrogation	1,2, 1, 2	3, 1,2, 1,2,3	2, 1, 1	1,4 1	–
Friendly Face (Murali et al. 2021)	Public speaking	2, 2	–	1, 2, 1	1, 1	2, 2, 1,3,4
Nakash et al. (2022)	Suicide prevention	1, 2	–	–	1, 1	–
Communicate! (Jeuring et al. 2015)	Communication Skills	3, 1	1	2, 2, 1	1,2, 1	–
Schoenthaler et al. (2017), Albright et al. (2018)	Medical communication	3, 1, 1	2, 1,2, 1,2	1,2, 2, 1	1,2,3, 2	2, 1, 1
Maicher et al. (2017)	History taking	2, 2	1	–	1, 1	–
O'Brien et al. (2019)	Suicide prevention	2, 1, 1	2, 1,2, 1	1,2, 1,2, 1	2,4, 2	2, 1, 1,3
Muller et al. (2012)	Sales training	3, 1, 1	3, 1, 1,3	1, 1, 1	2,4, 1	–
Yao et al. (2022, 2020)	Medical empathy training	2, 2	1	2, 2, 1	2,4, 2	1, 1, 3
ELECT BiLAT (Lane et al. 2007)	Intercultural negotiation	2, 1,2, 1	2, 1,2, 1,2,3	1,2, 1,2, 1	1,2,4, 2	2, 1,2, 1,2,3
Szilas et al. (2019)	Caregivers training	3, 1	1	–	1, 1	–
Ochs et al. (2019)	Medical communication	1,2, 2	–	2, 2, 1	1,4, 1	–
Georgila et al. (2019)	Counseling skills	3, 1	1	1,2, 2, 1	1,2, 2	1, 1, 1,3
Ziebarth et al. (2014)	Medical Interviews	2, 1, 2	2, 1,2, 1	1,2, 1,2, 1,2	1,2,4, 2	–
Adewole et al. (2020)	Cultural Awareness	3, 2	1	1, 2, 1	2,4 2	1, 1, 1,3
Bánszki et al. (2018)	Medical education	1, 2	–	–	1, 1	–
UT TIME Portal (Zielke et al. 2016)	Medical communication	3, 1	1	1,2, 1,2, 1	1,2,3, 2	–
Zlotos et al. (2016)	Medical communication	2, 1	1	–	1, 1	–
AdaCoach (Peng et al. 2022)	Customer service	3, 2	1	1, 2, 1	1, 1	–
Dupuy et al. (2020)	Medical interviews	3, 2	1	1, 2, 1	1, 1	–
Mission rehearsal exercise (Hill et al. 2003)	Decision-making skills	3, 1, 2	3, 1,2, 1,2,3	–	1,4, 1	–
Peddle et al. (2019)	Medical communication	3, 1	1	–	1, 1	–
deLearyous (Vaassen et al. 2012)	Communication skills	2, 1, 2	3, 1, 2	–	1, 1	–
Washburn, Parrish, and Bordnick (2020)	Medical education	2, 2	1	2, 2, 1	1,4, 1	–
Demasi, Li and Yu (2020)	Hotline counselling	3, 1,2, 2	1	–	1,4, 1	–
INOTS (Hays et al. 2012)	Interpersonal skills	3, 2	1	2, 2, 1	2,3, 1	–
Virtual-Suspect (Bitan et al. 2017)	Suspect interrogation	2, 1, 2	2, 1, 1	–	1,4, 1	–
Mell and Gratch (2017), Johnson et al. (2019)	Negotiations	2, 1,2, 1,2	3, 1,2, 1,3	2, 2, 1	1,2,4, 2	–
Jacklin, Maskrey, and Chapman (2018), Jacklin, Chapman, and Maskrey (2019)	Shared decision making	3, 1	1	2, 2, 1	1,2, 1	–

(*Continued*)

Table 6. Continued.

System	Training type	Examined components: Action decision making	Agent model	Feedback	Instructional principles	Pedagogical agent
Sveinbjörnsdóttir et al. (2019)	Training teachers	2, 1	1	1, 2, 1	2,4, 2	–
DialogueTrainer (n.d.)[a]	Several skills	3, 1	1	2, 2, 1,2	1,2, 1	2, 2, 2
SIMmersion (n.d.)[a]	Several skills	2, 1	1	1,2, 1,2, 1	1,2,4, 2	1, 1, 1
Kognito (n.d.)[a]	Several skills	3, 1	1	1,2, 1,2, 1	1,2,4, 2	2, 1, 1
Mursion (n.d.)[a]	Several skills	1, 2	–	–[b]	1,2,4, 2	2, 2, 4
VirtualSpeech (n.d.)[a]	Several skills	2, 2	1	1,2, 2, 1,2	1,2, 1	–
CleVR (n.d.)[a]	Several skills	1, 2	–	–	1,4, 1	–

(a) These systems are commercially available, and were categorised based on information from their websites and publicly available data. Thus, some information might be missing from the categorisation. (b) A human actor gives feedback.

by Cognitivism (60%), Constructivism (49%), and Social Cognitive Theory (12%). This prevalence might stem from training systems' aim to train learners on social situation responses, often focusing on behaviour formation. Lastly, a mere 28% of the systems featured a pedagogical agent, either as text (9%) or an embodied agent (19%).

The table also shows that all five components are present in the examined training systems, suggesting their relevance in such contexts. Each concept discussed in the categorisation table (Table 5) also appears in at least two of the examined systems, further validating the relevance of the ideas. Figure 6 shows the extent to which the 43 systems cover the agent simulation model and the educational model. We assessed the agent simulation based on the complexity of decision-making and the agent model, and we assessed the educational model based on the presence of three elements: feedback, multiple instructional principles, and the use of a pedagogical agent. None of the systems achieved the highest scores in both dimensions, while some scored highly in mainly a single dimension, suggesting a focus on either one over the other.[2] Furthermore, the training systems analysed span various topics, from job interviews to suicide prevention and negotiations. This broad representation emphasises the general applicability of the five components across various training types, thus supporting ARTES's generalisability.

4.3. Tools support for ARTES components

Our previous discussion examined the link between ARTES components and other architectures conceptually. Now, we turn our attention to the implementation support for ARTES components, i.e. do existing

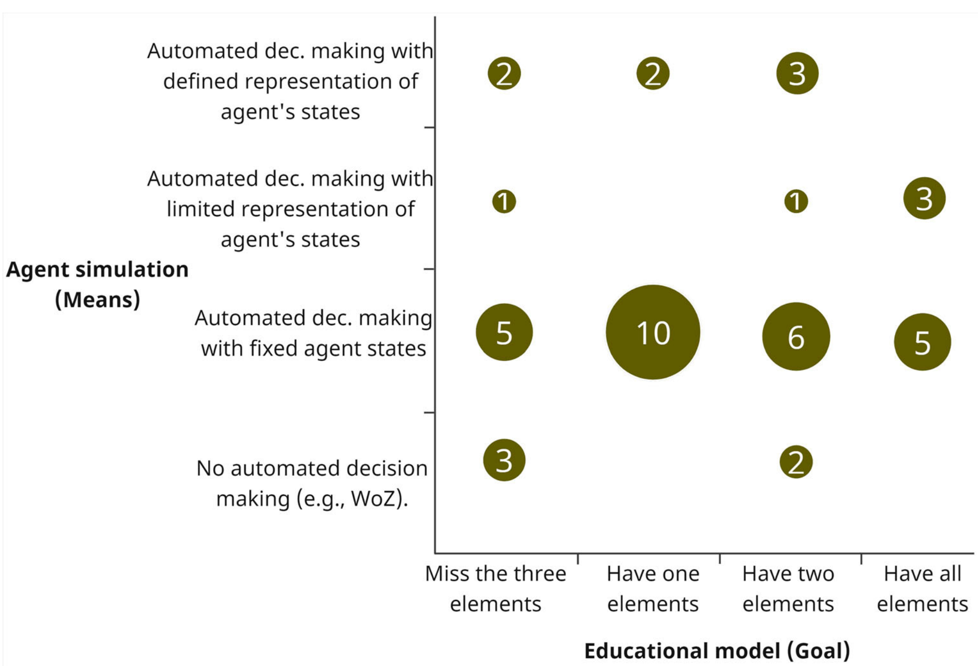


Figure 6. Categorisation of the 43 training systems based on criteria related to the agent simulation model and the educational model, as detailed in Tables 5 and 6. The numbers inside the circles represent the count of systems that correspond to each category.

Table 7. Mapping virtual human platforms and their tools to ARTES's agent simulation components.

ARTES components	GRETA	VHToolkit	Agents United
Simulation environment	FAP-BAP Player Ogre3D Unity Robotics	Unity	Unity Robotics
Speech recognition	PureData	MS SAPI Google Voice PocketSphinx	N/A
Multimodal perception and understanding	SSI OpenFace OpenSmile	MultiSense	Holistic Behaviour Analysis Framework AWARE universAAL
NLU	Emotional Mind Dialog Manager	NPCEditor	WOOL
Agent model & Action decision making	Intent Planner TRINDIKIT MIDAS	NPCEditor Scripting	Topic Selection Engine Dialogue Game Execution Flipper WOOL
NLG	Emotional Mind Dialog Manager	NPCEditor	WOOL
Speech generation	MaryTTS CereVoice	Festival MS SAPI Rhetorical CereVoice	MaryTTS MS SAPI
Nonverbal behaviour generation	Behavior Planner	NVBG	Flipper
Nonverbal behaviour realizer	Behavior Realizer	SmartBody Unity BML Realizer	GRETA Behavior Realizer ASAP Realizer
Animation system	OpenGL Library Unity Mecanim	Unity Mecanim	Unity Mecanim

tools support the implementation of ARTES components? To answer this question, we linked ARTES to three development platforms for creating virtual humans: GRETA (Poggi et al. 2005), the Virtual Human Toolkit (VHToolkit) (Hartholt et al. 2022, 2013), and Agents United (Beinema et al. 2021). While these platforms are mainly designed to develop virtual humans, they are also suitable for agent-based social skills training systems. Table 7 depicts which tools can be utilized to build ARTES agent simulation components in these three platforms. The linking shows that the tools can support the implementation of ARTES simulation components. For a comprehensive overview of virtual human tools and their applications, we refer to the chapter by Hartholt and Mozgai (2022).

Regarding ARTES's educational model, we can map ITS-based authoring tools directly to it, as ARTES's educational model also relies on ITS main components (domain knowledge, learner knowledge, and tutor). ITS-centric examples include GIFT (Sottilare et al. 2017) and AutoTutor (Graesser et al. 2004). Integrating authoring tools with virtual human platforms would align with the two main parts of ARTES: the virtual agent simulation and the ITS-based educational model. Such integration is possible, as shown by merging the VHToolkit with AutoTutor (Swartout et al. 2016), and the VHToolkit with GIFT (Brawner, Hoffman, and Nye 2019). In the future, ARTES could be evaluated through an example implementation using such tools or by implementing a prototype from scratch, allowing us to evaluate other quality attributes such as performance and usability (Abowd et al. 1997).

5. Further research directions

What research directions still warrant further exploration? We identified nine of them, which we will discuss here one by one.

Knowledge Transfer. Knowledge transfer in education relates to learners' ability to generalise the gained knowledge to other contexts (Barnett and Ceci 2002). In training systems, this translates to transferring knowledge gained from agent simulations to real-world situations. Various concerns arise regarding knowledge transfer in training systems, notably: (a) the inaccuracy of reflecting complex, real-world interactions; (b) the simulated situation being context-dependent; and (c) the lack of motivation to learn and transfer skills to other contexts. Despite the transfer's importance, most examined systems overlooked the assessment of real-life knowledge. In their review, Bosman, Bosse and Formolo (2018) reported one system out of twelve that tested for knowledge transfer, attributed to the cost and expertise needed to conduct it, especially as computer scientists in academia mainly created these systems. Naturally, testing for such tangible outcomes should accompany adherence to defined design guidelines and recommendations for fostering knowledge transfer (e.g. Craig and Schroeder 2018). Regarding the provided feedback, away from the agent simulation, it is worth noting the critical role of instructions specificity level for knowledge transfer. Eiriksdottir and Catrambone (2011), for instance, examined educational instructions and found that participants receiving detailed, step-by-step instructions performed well initially but received worse learning and transfer results later. On the other hand, they found that general instructions aided people in establishing a more adaptable reference framework for new tasks. This finding supports the need for knowledge transfer testing in agent-based training systems, as good initial performance results may not necessarily translate into successful knowledge transfer.

Balance between realism and educational gains. High-fidelity simulation does not necessarily correspond to better learning outcomes (Carey and Rossler 2020). For training systems, the realism of a simulation could cover elements such as the rendered agent and environment, the agent's behaviour, dialogues, and sound. Gallagher et al. (2005) argued that the essential question is whether the simulation trains the appropriate skills, regardless of the technological fidelity. At the same time, realism in simulation can impact attributes such as the agent's likability (Baylor and Kim 2004) and the satisfaction of the agent (Prajod et al. 2019), which could, in turn, influence educational gains. Thus, the challenge lies in identifying essential simulated elements that can fulfil learning goals. More specifically, how to find cues or behaviours that are sufficiently realistic to train specific skills and meet the training objectives. For instance, adjusting eye gaze intensity can help manage social phobia in virtual reality exposure therapy (Brinkman et al. 2012), or altering depth perception can address fear of heights (Ling et al. 2013). As for establishing a balance between training objectives and simulation fidelity, we took steps towards it by linking learning theories to ARTES components, which would contribute to determining the level of development required for each component, guided by the targeted skills and relevant theories.

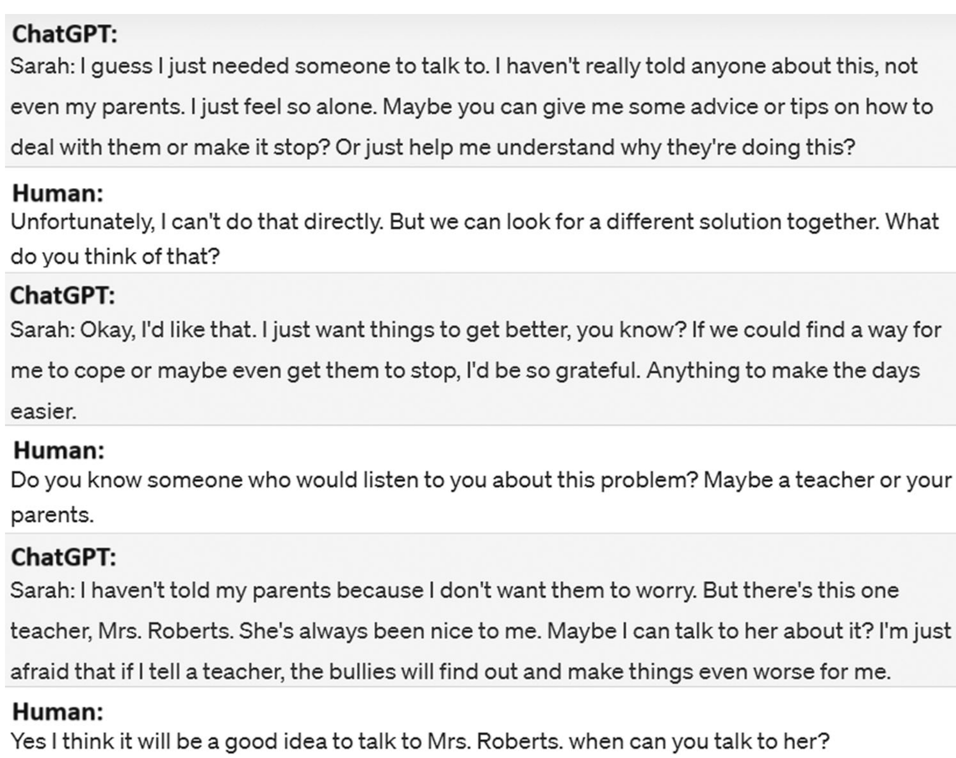
Explainablity of training systems decisions. The nature of training systems requires two types of explanations for learners: one addressing the virtual agent's behaviour, and the other offering insights into the learner's choices as part of the educational process. There is a need to integrate these two aspects within the context of training systems. Doing that would help generate and present reliable and valid feedback that explains the agent's actions and the learner's decisions. When it comes to the explainability of the agent, we need to explain the agent's actions, their consequences, and why the scenario unfolded in the way it did. This is often referred to as explainable agency, and education is one of its drives for demanding explainability (Anjomshoae et al. 2019). This explainability typically unfolds in three phases: generating the explanation regarding the action, communicating the reason to the learner, and assessing the explanation reception (Neerincx et al. 2018)- the latter equivalent to experiments in the case of training systems. The other type of explanation is regarding the learner's performance in training. One such work in this domain is where Khosravi et al. (2022) defined a framework for Explainable AI in education, which are six dimensions that education systems developers and researchers should consider.

Ethical implications. Using interactive agents in training systems could raise stereotypes and biases. Rivera-Gutierrez et al. (2014) experimented with six virtual patients varying in gender and skin colour, revealing a disparity in correct diagnoses among the different characters. The authors speculated this could be due to the transfer of prejudices from real life to the simulation. Conversely, this transfer could potentially develop biases based on the learner experience within the training system. For instance, if a character with certain traits behaves in a particular way, it could inadvertently reinforce discriminatory attitudes towards such traits or populations. Moreover, modelling 'appropriate' behaviour in agent-based training systems raises two main concerns: regarding (1) inaccurate design and (2) potential malicious control of this behaviour. If an agent's behaviour is poorly designed, simulations might not accurately represent human interactions. This could lead to ineffective training, giving learners a false sense of competence. As for who is controlling appropriate behaviour, training systems could be used to shape and moralise people on specific behaviours. Similarly, ill-intentioned learners could misuse the skills learned from training systems to cause harm, such as exacerbating bullying. Regarding affective privacy in virtual agents, Hudlicka (2016) highlighted three ethical concerns, which are apparent in a training system that tries to change the user's emotions, e.g. social anxiety training. The concerns are affective privacy, or the right to keep one's emotions private; emotion induction, the act of evoking and changing specific emotions in a person; and the formation of virtual relationships through (over)trust and attachment. If these systems are inadequately designed, for instance, by evoking inappropriate emotions in the learners, they could have negative consequences that exacerbate the issues.

Other ethical questions for training systems could arise while analysing and storing personal data and regarding consent and autonomy for such simulation training. A solution that could address this is to develop guidelines and checklists as a community, which can provide a structure for design validation. Furthermore, we should involve multiple stakeholders in the evaluation of training systems to get a variety of perspectives on system validity and preferences (e.g. Botezatu et al. 2010), and by letting them participate in the design process (e.g. Kleinsmith et al. 2015; Kron et al. 2017).

Large Language Models (LLMs). These models, such as GPT (Brown et al. 2020), can be considered as another option for simulating a person in a training scenario. Figure 7 shows an interaction with ChatGPT acting as a child contacting a children's helpline. In this case, the conversations appear convincing and reasonably coherent. However, there are currently several challenges to the direct application of LLM in social skills training systems. For example, it is less clear how to explain agents' cognitions in the absence of a human-understandable agent model. Additionally, LLMs are reported to suffer from bias issues (Nadeem, Bethke and Reddy 2020) and carry a risk of misinformation in the agent reply (Weidinger et al. 2021), which could deviate the training from its objectives. One possible approach to utilizing LLMs is to simulate some parts of the agent's thinking. These models generally generate fitting responses, but may occasionally deliver incorrect information (Kasneci et al. 2023). Therefore, exploring hybrid solutions could be beneficial, blending rule-based agent models and LLMs to generate appropriate agent responses. This could involve generating parts of the agent's thinking, like their beliefs, intentions, and actions, to be fed into a framework (Antunes et al. 2023). Another outlook involves incorporating LLMs into the tutor model to supplement or potentially replace certain components. Despite their limitations, LLMs could automate sections of the educational process and reply to unexpected learner inputs. For Lila's case, three examples can illustrate the potential applications of LLM-based tutor subcomponents. First, a pedagogical agent could serve as a social actor, addressing Lila's questions about communication protocols and counselling in general. Second, the responses of the virtual child character could be personalised based on Lila's demonstrated proficiency, thus reflecting instructional principles. Lastly, feedback for Lila could be generated by comparing her conversation logs with the virtual child to an ideal conversation log.

Crowdsourcing solutions. As previously mentioned, data required to create a virtual human or a tutoring


ChatGPT:
Sarah: I guess I just needed someone to talk to. I haven't really told anyone about this, not even my parents. I just feel so alone. Maybe you can give me some advice or tips on how to deal with them or make it stop? Or just help me understand why they're doing this?

Human:
Unfortunately, I can't do that directly. But we can look for a different solution together. What do you think of that?

ChatGPT:
Sarah: Okay, I'd like that. I just want things to get better, you know? If we could find a way for me to cope or maybe even get them to stop, I'd be so grateful. Anything to make the days easier.

Human:
Do you know someone who would listen to you about this problem? Maybe a teacher or your parents.

ChatGPT:
Sarah: I haven't told my parents because I don't want them to worry. But there's this one teacher, Mrs. Roberts. She's always been nice to me. Maybe I can talk to her about it? I'm just afraid that if I tell a teacher, the bullies will find out and make things even worse for me.

Human:
Yes I think it will be a good idea to talk to Mrs. Roberts. when can you talk to her?

Figure 7. A chat from ChatGPT (GPT-4). We prompt the tool to act as a bullied child contacting a children's helpline. The prompt was '*Engage in roleplay. You are a child (Sarah) who was bullied at school and chatting with a child helpline. I will play the volunteer at the helpline. Let's start: "I am John, I am here to listen to you. What do you like to talk about?"*'.

model is not always accessible or adequately expansive for every training system domain. One solution to such a challenge is crowdsourcing. In theory, crowdsourcing could aid in collecting data required for a data-driven virtual agent. However, in most training systems, a domain expert is needed to feed and evaluate such data, thereby limiting the pool of workers. Nevertheless, crowdsourcing has proven helpful for training systems; for example, Rossen and Lok (2012) proposed an approach to develop a conversation with a virtual agent rapidly. This approach allows a domain expert to start an agent with an initial set of questions and responses, which will then interact with novices recruited through crowdsourcing. The domain expert then refines the agent with information from novice interactions. This iterative procedure continues until the agent's performance reaches a satisfactory level. Such an approach could decrease the cognitive load and workload on the domain expert and yield a broader range of responses. Another approach that focuses on simulating different actions in a situation was proposed by Feng et al. (2018). In that study, the workers were asked to complete a situation story and how it could unfold. These stories were fed to data-driven models, and the workers evaluated the outputs iteratively. Although this approach reveals actions and consequences within a situation, it falls short in modelling conversations – meaning the agents' dialogues are not directly trained. Still, it can be an initial step to creating a good story for future dialogues. On the agent model level, crowd workers could contribute to defining the agent's states, based on the current state of the conversation.

Multi-agent scenarios. We, with ARTES, only focused on social scenarios involving a single virtual agent. However, training social skills might require the presence of multiple actors, and thus, there is a need to model more than one agent to simulate a situation. Depending on a scenario's context and learning objectives, agents could have individual cognition or share a collective cognitive model. For example, a virtual nurse and a virtual patient could be simulated to train a student doctor (Kron et al. 2017). Furthermore, multiple agents can be used with social cognitive theory to model agents with differing perspectives on a topic that a learner could learn from. For example, learners can explore the ethical implications of implementing a new two-factor authentication policy by engaging with agents that prioritize different ethical principles, such as non-maleficence (not harming people) or autonomy (giving people choice) (Ali Bajwa, Richards, and Formosa 2023). It would be interesting to explore the dynamics of multiple agents and how their shared cognition could enhance the learner's training. Taking this concept further, the simulation system could train more than one learner in a collaborative learning environment. An example of this approach is utilising a virtual agent moderating conflict resolution between two learners (Emmerich et al. 2012).

Research methods. Prior research has discussed the lack of standardised terminology and standardised evaluation of learning outcomes for virtual patients (Battegazzorre, Bottino, and Lamberti 2020) and training systems in general (Bosman, Bosse and Formolo 2018). As a next step, the community could establish strategies to enhance the reproducibility of results. There are two strategies to foster the future reproducibility of the system's positive learning effects. The first one focuses on specifying the material used, for example, a model to illustrate the internal functioning of the agent. However, crafting an accurate model presents its challenges, as these models represent an abstraction of reality. Thus, an alternative strategy is to record people's impressions and test whether these can be replicated as a blackbox approach. One of the initiatives for this strategy is the Artificial Social Agents (ASA) questionnaire (Fitrianie et al. 2022), which allows researchers to measure a broad number of interaction impressions, such as the believability, enjoyability, and social presence of an agent (Fitrianie et al. 2021).

Procedurally generated scenarios. Regarding social interaction training systems, there are many intersections with serious games – games designed for objectives beyond entertainment. They both have goals, objectives, and, usually, virtual characters. Incorporating gamification elements can increase engagement and overall learning outcomes (Huang et al. 2020; Sailer and Homner 2020). This can be achieved by using scores as rewards (Gebhard et al. 2018; Jeuring et al. 2015), or using badges (Zielke et al. 2016). One direction forward for training systems is incorporating procedurally generated dialogue, where the storyline, characters and their proprieties (e.g. traits and goals) dynamically evolve as the scenario progresses. For example, the game 'Tech Support: Error Unknown'[3], which, despite not being the main focus, does exhibit some training elements. Here, the player acts as a technical support representative who people contact for support, where the customer's support requests are procedurally generated. Having a similar procedural narrative could introduce an element of randomness to scenarios in training systems, thus mitigating repetitiveness and predictability.

6. Final remarks

We believe ARTES is a valuable foundation for researchers and developers in agent-based social skills training systems. This paper's contribution lies in the proposal of this versatile architecture and the examination of its key components, comparing them to similar domains and linking them to learning theories. Understanding the complexities and roles of these vital components, we hope, could enhance the efficiency, cost-effectiveness, adaptability, and impact of training systems. Therefore, our work serves as a stepping-stone to bridging the gap between virtual human simulations and effective educational learning.

Notes

1. Note that the mapping does not evaluate the quality or implementations of the architectures, but rather showcases ARTES's coverage. These architectures may support additional functionalities and libraries beyond what is reported in the examined literature.
2. Similar to the mapping of architectures, systems are typically designed to suit their specific contexts. Thus, our analysis does not assess their overall quality; instead, it aims to show how these systems incorporated ARTES's components, highlighting its relevance and modularity of the architecture.
3. Dragon Slumber, Tech support: Error unknown, [Steam], 2019.

Acknowledgments

We want to thank Tibor Bosse, Jordy van Dortmont, Niels ter Heijden, Thijs Hendrickx, Carolina Jorge, Esra de Groot, Mani Tajaddini, Michaël Grauwde, Pei-Yu Chen and Siddharth Mehrotra for their valuable contributions to this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by King Saud University and the Saudi Arabian Cultural Mission (SACM).

Data availability statement

The analysis of the double-coding is available through the 4TU research data repository: https://data.4tu.nl/datasets/a037707b-5d70-410a-b6e1-9ee4a694b1d7

ORCID

Mohammed Al Owayyed http://orcid.org/0000-0002-9680-9204
Myrthe Tielman http://orcid.org/0000-0002-7826-5821
Arno Hartholt http://orcid.org/0000-0001-7950-0048
Marcus Specht http://orcid.org/0000-0002-6086-8480
Willem-Paul Brinkman http://orcid.org/0000-0001-8485-7092

References

Abowd, G., L. Bass, P. Clements, R. Kazman, L. Northrop, and A. Zaremski. 1997. "Recommended Best Industrial Practice for Software Architecture Evaluation." Technical Report CMU/SEI-96-TR 25.

Adam, C., and B. Gaudou. 2016. "BDI Agents in Social Simulations: A Survey." *The Knowledge Engineering Review* 31 (3): 207–238. https://doi.org/10.1017/S0269888916000096.

Adewole, S., E. Gharavi, B. Shpringer, M. Bolger, V. Sharma, S. M. Yang, and D. E. Brown. 2020. "Dialogue-Based Simulation for Cultural Awareness Training. arXiv preprint arXiv:2002.00223.

Akyuz, Y. 2020. "Effects of Intelligent Tutoring Systems (ITS) on Personalized Learning (PL)." *Creative Education* 11 (6): 953–978. https://doi.org/10.4236/ce.2020.116069.

Albright, G., C. Bryan, C. Adam, J. McMillan, and K. Shockley. 2018. "Using Virtual Patient Simulations to Prepare Primary Health Care Professionals to Conduct Substance Use and Mental Health Screening and Brief Intervention." *Journal of the American Psychiatric Nurses Association* 24 (3): 247–259. https://doi.org/10.1177/1078390317719321.

Ali Bajwa, M. H., D. Richards, and P. Formosa. 2023. "Evaluation of Embodied Conversational Agents Designed with Ethical Principles and Personality for Cybersecurity Ethics Training." In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*, 1–8. Würzburg: Association for Computing Machinery, Inc.

Alkhatlan, A., and J. Kalita. 2018. "Intelligent Tutoring Systems: A Comprehensive Historical Survey with Recent Developments." arXiv preprint arXiv:1812.09628.

Anderson, K., E. André, T. Baur, S. Bernardini, M. Chollet, E. Chryssafidou, I. Damian, et al. 2013. "The Tardis Framework: Intelligent Virtual Agents for Social Coaching in Job Interviews." In *International Conference on Advances in Computer Entertainment Technology*, 476–491. Springer.

Anderson, J. R., C. F. Boyle, and B. J. Reiser. 1985. "Intelligent Tutoring Systems." *Science* 228 (4698): 456–462. https://doi.org/10.1126/science.228.4698.456.

Anjomshoae, S., A. Najjar, D. Calvaresi, and K. Främling. 2019. "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*, 1078–1088. International Foundation for Autonomous Agents and Multiagent Systems.

Antunes, A., J. Campos, M. Guimarães, J. Dias, and P. A. Santos. 2023. "Prompting for Socially Intelligent Agents with ChatGPT." In *Proceedings of the 23nd ACM International Conference on Intelligent Virtual Agents*. Würzburg: Association for Computing Machinery, Inc.

Aylett, R. S., S. Louchart, J. Dias, A. Paiva, and M. Vala. 2005. "Fearnot! – An Experiment in Emergent Narrative." In *Intelligent Virtual Agents: 5th International Working*

Conference, IVA 2005, Kos Greece, September 12–14, 2005, 305–316. Springer.

Bánszki, F., J. Beilby, M. Quail, P. Allen, S. Brundage, and J. Spitalnick. 2018. "A Clinical Educator's Experience Using a Virtual Patient to Teach Communication and Interpersonal Skills." *Australasian Journal of Educational Technology* 34 (3): 60–73.

Baptista, M. L., C. R. Martinho, F. Lima, P. A. Santos, and H. Prendinger. 2014. "An Agent-Based Model of Consumer Behavior based on the BDI Architecture and Neoclassical Theory." In *Developments in Business Simulation and Experiential Learning: Proceedings of the Annual ABSEL Conference*. Vol. 41. Paris: Springer Verlag.

Barnett, S. M., and S. J. Ceci. 2002. "When and Where Do We Apply what We Learn?: A Taxonomy for Far Transfer." *Psychological Bulletin* 128 (4): 612–637. https://doi.org/10.1037/0033-2909.128.4.612.

Battegazzorre, E., A. Bottino, and F. Lamberti. 2020. "Training Medical Communication Skills with Virtual Patients: Literature Review and Directions for Future Research." In *International Conference on Intelligent Technologies for Interactive Entertainment*, 207–226. Springer.

Baylor, A. L., and Y. Kim. 2004. "Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity and Instructional Role." In *International Conference on Intelligent Tutoring Systems*, 592–603. Springer.

Beidel, D. C., C. A. Alfano, M. J. Kofler, P. A. Rao, L. Scharfstein, and N. W. Sarver. 2014. "The Impact of Social Skills Training for Social Anxiety Disorder: A Randomized Controlled Trial." *Journal of Anxiety Disorders* 28 (8): 908–918. https://doi.org/10.1016/j.janxdis.2014.09.016.

Beinema, T., D. Davison, D. Reidsma, O. Banos, M. Bruijnes, B. Donval, Á. F. Valero, et al. 2021. "Agents United: An Open Platform for Multi-Agent Conversational Systems." In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*, Virtual Event, Japan, 17–24. Association for Computing Machinery, Inc.

Bitan, M., G. Nahari, Z. Nisin, A. Roth, and S. Kraus. 2017. "Psychologically based Virtual-Suspect for Interrogative Interview Training." In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31. San Francisco: AAAI press.

Bosman, K., T. Bosse, and D. Formolo. 2018. "Virtual Agents for Professional Social Skills Training: An Overview of the State-of-the-Art." In *International Conference on Intelligent Technologies for Interactive Entertainment*, 75–84. Springer.

Bosse, T., J. De Man, and C. Gerritsen. 2014. "Agent-Based Simulation as a Tool for the Design of a Virtual Training Environment." In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI), Intelligent Agent Technologies (IAT)*,'Vol. 3, 40–47. IEEE.

Bosse, T., C. Gerritsen, and J. de Man. 2016. "An Intelligent System for Aggression De-Escalation Training." In *ECAI 2016*, 1805–1811. IOS Press.

Botezatu, M., H. Hult, M. Kassaye Tessma, and U. G. Fors. 2010. "As Time Goes by: Stakeholder Opinions on the Implementation and Use of a Virtual Patient Simulation System." *Medical Teacher* 32 (11): e509–e516. https://doi.org/10.3109/0142159X.2010.519066.

Bourdeau, J., and M. Grandbastien. 2010. "Modeling Tutoring Knowledge." In *Advances in Intelligent Tutoring Systems*, 123–143. Springer.

Bourgais, M., P. Taillandier, and L. Vercouter. 2019. "BEN: An Agent Architecture for Explainable and Expressive Behavior in Social Simulation." In *Explainable, Transparent Autonomous Agents and Multi-Agent Systems: First International Workshop, EXTRAAMAS 2019*, Montreal, QC, Canada, May 13–14, 2019, 147–163. Springer. Revised Selected Papers 1.

Brabra, H., M. Báez, B. Benatallah, W. Gaaloul, S. Bouguelia, and S. Zamanirad. 2021. "Dialogue Management in Conversational Systems: A Review of Approaches, Challenges, and Opportunities." *IEEE Transactions on Cognitive and Developmental Systems* 14 (3): 783–798.

Brawner, K., M. Hoffman, and B. Nye. 2019. "Architecture, Ontology in the Generalized Intelligent Framework for Tutoring: 2018 Update." In *7th Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium*, 11.

Brinkman, W.-P., D. Hartanto, N. Kang, D. de Vliegher, I. L. Kampmann, N. Morina, P. G. Emmelkamp, and M. Neerincx. 2012. "A Virtual Reality Dialogue System for the Treatment of Social Phobia." In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, 1099–1102.

Brown, T., B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan. 2020. "Language Models are Few-Shot Learners." *Advances in Neural Information Processing Systems* 33:1877–1901.

Bruijnes, M. 2016. Believable Suspect Agents: Response and Interpersonal Style Selection for an Artificial Suspect.

Bruijnes, M., R. op den Akker, A. Hartholt, and D. Heylen. 2015. "Virtual Suspect William." In *Intelligent Virtual Agents: 15th International Conference, IVA 2015*, Delft, The Netherlands, August 26–28, 2015, 67–76. Springer.

Carey, J. M., and K. Rossler. 2020. The How When Why of High Fidelity Simulation.

Cheng, Y.-M., L.-S. Chen, H.-C. Huang, S.-F. Weng, Y.-G. Chen, and C.-H. Lin. 2009. "Building a General Purpose Pedagogical Agent in a Web-Based Multimedia Clinical Simulation System for Medical Education." *IEEE Transactions on Learning Technologies* 2:216–225. https://doi.org/10.1109/TLT.2009.18.

Clark, R. E. 2007. "Learning From Serious Games? Arguments, Evidence, and Research Suggestions." *Educational Technology* 47:56–59.

Clements, P. C., and L. M. Northrop. 1996. "Software Architecture: An Executive Overview." Carnegie Mellon University, Software Engineering Institute.

CleVR. n.d. https://clevr.net/.

Core, M., H. C. Lane, and D. Traum. 2014. "Intelligent Tutoring Support for Learners Interacting with Virtual Humans." *Design Recommendations for Intelligent Tutoring Systems* 2:249–257.

Craig, S. D., and N. L. Schroeder. 2018. "Design Principles for Virtual Humans in Educational Technology Environments." In *Deep Comprehension*, 128–139. Routledge.

de Beyn, A. 2003. IN GESPREK MET KINDEREN: de methodiek van de Kindertelefoon. NIZW.

Demasi, O., Y. Li, and Z. Yu. 2020. A Multi-Persona Chatbot for Hotline Counselor Training." In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 3623–3636. Association for Computational Linguistics.

DialogueTrainer. n.d. https://www.dialoguetrainer.com.

Dias, J., S. Mascarenhas, and A. Paiva. 2014. "Fatima Modular: Towards an Agent Architecture with a Generic Appraisal Framework." In *Emotion Modeling*, 44–56. Springer.

Ding, D. 2020. "Design and Evaluation of Simulated Reflective Thoughts in Virtual Reality Exposure Training." PhD thesis, Delft University of Technology.

Doberstein, D., N. Agreiter, M. Bäumer, M. Cui, S. Abdollahzadegan, D. Heidari, N. Jiang, et al. 2016. "Cucomag-Group Reflection Support in Role-Playing Environments." DeLFI 2016–Die 14. E-Learning Fachtagung Informatik.

Dragon Slumber 2019. Tech support: Error unknown, [Steam].

Dupuy, L., J.-A. Micoulaud-Franchi, H. Cassoudesalle, O. Ballot, P. Dehail, B. Aouizerate, E. Cuny, E. de Sevin, and P. Philip. 2020. "Evaluation of a Virtual Agent to Train Medical Students Conducting Psychiatric Interviews for Diagnosing Major Depressive Disorders." *Journal of Affective Disorders* 263:1–8. https://doi.org/10.1016/j.jad.2019.11.117.

Eiriksdottir, E., and R. Catrambone. 2011. "Procedural Instructions, Principles, and Examples: How to Structure Instructions for Procedural Tasks to Enhance Performance, Learning, and Transfer." *Human Factors* 53 (6): 749–770. https://doi.org/10.1177/0018720811419154.

Emmerich, K., K. Neuwald, J. Othlinghaus, S. Ziebarth, and H. U. Hoppe. 2012. "Training Conflict Management in a Collaborative Virtual Environment." In *International Conference on Collaboration and Technology*, 17–32. Springer.

Ertmer, P. A., and T. J. Newby. 1993. "Behaviorism, Cognitivism, Constructivism: Comparing Critical Features From An Instructional Design Perspective." *Performance Improvement Quarterly* 6 (4): 50–72. https://doi.org/10.1111/piq.1993.6.issue-4.

Feine, J., U. Gnewuch, S. Morana, and A. Maedche. 2019. "A Taxonomy of Social Cues for Conversational Agents." *International Journal of Human-Computer Studies* 132:138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009.

Feng, D., P. Sequeira, E. Carstensdottir, M. S. El-Nasr, and S. Marsella. 2018. Learning Generative Models of Social Interactions with Humans-in-the-Loop." In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 509–516. IEEE.

Fitrianie, S., M. Bruijnes, F. Li, A. Abdulrahman, and W.-P. Brinkman. 2022. "The Artificial-Social-Agent Questionnaire: Establishing the Long and Short Questionnaire Versions." In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*, 1–8. Faro: Association for Computing Machinery, Inc.

Fitrianie, S., M. Bruijnes, F. Li, and W.-P. Brinkman. 2021. "Questionnaire Items for Evaluating Artificial Social Agents-Expert Generated, Content Validated and Reliability Analysed." In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents, Virtual Event, Japan*, 84–86. Association for Computing Machinery, Inc.

Gallagher, A. G., E. M. Ritter, H. Champion, G. Higgins, M. P. Fried, G. Moses, C. D. Smith, and R. M. Satava. 2005. "Virtual Reality Simulation for the Operating Room: Proficiency-Based Training As a Paradigm Shift in Surgical Skills Training." *Annals of Surgery* 241 (2): 364–372. https://doi.org/10.1097/01.sla.0000151982.85062.80.

Gebhard, P., T. Schneeberger, E. André, T. Baur, I. Damian, G. Mehlmann, C. König, and M. Langer. 2018. "Serious Games for Training Social Skills in Job Interviews." *IEEE Transactions on Games* 11:340–351. https://doi.org/10.1109/TGAMES.7782673.

Georgeff, M., B. Pell, M. Pollack, M. Tambe, and M. Wooldridge. 1998. The Belief-Desire-Intention Model of Agency." In *International Workshop on Agent Theories, Architectures, and Languages*, 1–10. Springer.

Georgila, K., M. G. Core, B. D. Nye, S. Karumbaiah, D. Auerbach, and M. Ram. 2019. Using Reinforcement Learning to Optimize the Policies of an Intelligent Tutoring System for Interpersonal Skills Training." In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 737–745. Montreal: International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS).

Graesser, A. C., S. Lu, G. T. Jackson, H. H. Mitchell, M. Ventura, A. Olney, and M. M. Louwerse. 2004. "Autotutor: A Tutor with Dialogue in Natural Language." *Behavior Research Methods, Instruments, & Computers* 36 (2): 180–192. https://doi.org/10.3758/BF03195563.

Gratch, J., D. DeVault, and G. Lucas. 2016. "The Benefits of Virtual Humans for Teaching Negotiation." In *International Conference on Intelligent Virtual Agents*, 283–294. Springer.

Guetterman, T. C., R. Sakakibara, S. Baireddy, F. W. Kron, M. W. Scerbo, J. F. Cleary, and M. D. Fetters. 2019. "Medical Students' Experiences and Outcomes Using a Virtual Human Simulation to Improve Communication Skills: Mixed Methods Study." *Journal of Medical Internet Research* 21:e15459. https://doi.org/10.2196/15459.

Harms, J.-G., P. Kucherbaev, A. Bozzon, and G.-J. Houben. 2018. "Approaches for Dialog Management in Conversational Agents." *IEEE Internet Computing* 23:13–22. https://doi.org/10.1109/MIC.2018.2881519.

Hartanto, D., I. L. Kampmann, N. Morina, P. G. Emmelkamp, M. A. Neerincx, and W.-P. Brinkman. 2014. "Controlling Social Stress in Virtual Reality Environments." *PLoS One* 9:e92804. https://doi.org/10.1371/journal.pone.0092804.

Hartholt, A., E. Fast, Z. Li, K. Kim, A. Leeds, and S. Mozgai. 2022. "Re-architecting the Virtual Human Toolkit: Towards an Interoperable Platform for Embodied Conversational Agent Research and Development." In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*, 1–8. Faro: Association for Computing Machinery, Inc.

Hartholt, A., and S. Mozgai. 2022. "Platforms and Tools for SIA Research and Development." In *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*. 1st ed., 261–304. New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3563659.3563668.

Hartholt, A., D. Traum, S. C. Marsella, A. Shapiro, G. Stratou, A. Leuski, L.-P. Morency, and J. Gratch. 2013. "All Together Now, Introducing the Virtual Human Toolkit." In *International Workshop on Intelligent Virtual Agents*, 368–381. Springer.

Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112. https://doi.org/10.3102/003465430298487.

Hays, M. J., J. C. Campbell, M. A. Trimmer, J. C. Poore, A. K. Webb, and T. K. King. 2012. "Can Role-Play with Virtual Humans Teach Interpersonal Skills." In *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*.

Hill, R. W., J. Gratch, S. Marsella, J. Rickel, W. R. Swartout, and D. R. Traum. 2003. "Virtual Humans in the Mission Rehearsal Exercise System." *Künstliche Intell.* 17 (4): 5.

Hirumi, A., T. Johnson, R. J. Reyes, B. Lok, K. Johnsen, D. J. Rivera-Gutierrez, K. Bogert. 2016. "Advancing Virtual Patient Simulations Through Design Research and Interplay: Part II – integration and Field Test." *Educational Technology Research and Development* 64 (6): 1301–1335. https://doi.org/10.1007/s11423-016-9461-6.

Hirumi, A., A. Kleinsmith, K. Johnsen, S. Kubovec, M. Eakins, K. Bogert, D. J. Rivera-Gutierrez, et al. 2016. "Advancing Virtual Patient Simulations Through Design Research and Interplay: Part I: Design and Development." *Educational Technology Research and Development* 64 (4): 763–785. https://doi.org/10.1007/s11423-016-9429-6.

Huang, R., A. D. Ritzhaupt, M. Sommer, J. Zhu, A. Stephen, N. Valle, J. Hampton, and J. Li. 2020. "The Impact of Gamification in Educational Settings on Student Learning Outcomes: A Meta-Analysis." *Educational Technology Research and Development* 68 (4): 1875–1901. https://doi.org/10.1007/s11423-020-09807-z.

Hudlicka, E. 2016. "Virtual Affective Agents and Therapeutic Games." In *Artificial Intelligence in Behavioral and Mental Health Care*, 81–115. Elsevier.

Jacklin, S., S. Chapman, and N. Maskrey. 2019. "Virtual Patient Educational Intervention for the Development of Shared Decision-Making Skills: A Pilot Study." *BMJ Simulation & Technology Enhanced Learning* 5 (4): 215–217. https://doi.org/10.1136/bmjstel-2018-000375.

Jacklin, S., N. Maskrey, and S. Chapman. 2018. "Improving Shared Decision Making Between Patients and Clinicians: Design and Development of a Virtual Patient Simulation Tool." *JMIR Medical Education* 4:e10088. https://doi.org/10.2196/10088.

Jackson, D., and A. Latham. 2022. "Talk to the Ghost: The Storybox Methodology for Faster Development of Storytelling Chatbots." *Expert Systems with Applications* 190:116223. https://doi.org/10.1016/j.eswa.2021.116223.

Jeuring, J., F. Grosfeld, B. Heeren, M. Hulsbergen, R. IJntema, V. Jonker, N. Mastenbroek, et al. 2015. "Communicate! – A Serious Game for Communication Skills –." In *Design for Teaching, Learning in a Networked World: 10th European Conference on Technology Enhanced Learning, EC-TEL 2015*, Toledo, Spain, September 15–18, 2015, 513–517. Springer.

Johnson, J., and A. Henderson. 2002. "Conceptual Models: Begin by Designing what to Design." *Interactions* 9:25–32. https://doi.org/10.1145/503355.503366.

Johnson, E., G. Lucas, P. Kim, and J. Gratch. 2019. "Intelligent Tutoring System for Negotiation Skills Training." In *Artificial Intelligence in Education: 20th International Conference, AIED 2019*, Chicago, IL, USA, June 25–29, 2019, 122–127. Springer.

Jonassen, D. H., and P. Henning. 1999. "Mental Models: Knowledge in the Head and Knowledge in the World." *Educational Technology* 39 (3): 37–42.

Kasneci, E., K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser. 2023. "Chatgpt for Good? On Opportunities and Challenges of Large Language Models for Education." *Learning and Individual Differences* 103:102274. https://doi.org/10.1016/j.lindif.2023.102274.

Kenny, P., A. Hartholt, J. Gratch, W. Swartout, D. Traum, S. Marsella, and D. Piepol. 2007. "Building Interactive Virtual Humans for Training Environments." In *Proceedings of I/ITSEC*. Vol. 174, 911–916.

Khosravi, H., S. B. Shum, G. Chen, C. Conati, Y.-S. Tsai, J. Kay, S. Knight, et al. 2022. "Explainable Artificial Intelligence in Education." *Computers and Education: Artificial Intelligence* 3:100074.

Kirschner, P., J. Sweller, and R. E. Clark. 2006. "Why Unguided Learning Does Not Work: An Analysis of the Failure of Discovery Learning, Problem-Based Learning, Experiential Learning and Inquiry-Based Learning." *Educational Psychologist* 41:75–86.

Kleinsmith, A., D. Rivera-Gutierrez, G. Finney, J. Cendan, and B. Lok. 2015. "Understanding Empathy Training with Virtual Patients." *Computers in Human Behavior* 52:151–158. https://doi.org/10.1016/j.chb.2015.05.033.

Kochmar, E., D. D. Vu, R. Belfer, V. Gupta, I. V. Serban, and J. Pineau. 2022. "Automated Data-Driven Generation of Personalized Pedagogical Interventions in Intelligent Tutoring Systems." *International Journal of Artificial Intelligence in Education* 32 (2): 323–349. https://doi.org/10.1007/s40593-021-00267-x.

Kognito. n.d. https://kognito.com/.

Korsgaard, C. M. 1996. *The Sources of Normativity*. Cambridge, UK: Cambridge University Press.

Kotseruba, I., and J. K. Tsotsos. 2020. "40 Years of Cognitive Architectures: Core Cognitive Abilities and Practical Applications." *Artificial Intelligence Review* 53 (1): 17–94. https://doi.org/10.1007/s10462-018-9646-y.

Kron, F. W., M. D. Fetters, M. W. Scerbo, C. B. White, M. L. Lypson, M. A. Padilla, G. A. Gliva-McConvey. 2017. "Using a Computer Simulation for Teaching Communication Skills: A Blinded Multisite Mixed Methods Randomized Controlled Trial." *Patient Education and Counseling* 100 (4): 748–759. https://doi.org/10.1016/j.pec.2016.10.024.

Laird, J. E. 2019. *The Soar Cognitive Architecture*. Cambridge, MA: MIT Press.

Landis, J. R., and G. G. Koch. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33 (1): 159–174. https://doi.org/10.2307/2529310.

Lane, H. C., M. G. Core, D. Gomboc, A. Karnavat, and M. Rosenberg. 2007. "Intelligent Tutoring for Interpersonal and Intercultural Skills." In Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC 2007), Orlando, FL.

Lee, J. 1999. "Effectiveness of Computer-Based Instructional Simulation: A Meta Analysis." *International Journal of Instructional Media* 26 (1): 71.

Lee, J., D. DeVault, S. Marsella, and D. Traum. 2008. "Thoughts on FML: Behavior Generation in the Virtual Human Communication Architecture." Why Conversational Agents do what they do. Functional Representations for Generating Conversational Agent Behavior, AAMAS 2008. Estoril.

Leslie, A. M. 1994. "ToMM, ToBy, and Agency: Core Architecture and Domain Specificity." In *Mapping the Mind: Domain Specificity in Cognition and Culture,* edited by Lawrence A. Hirschfeld and Susan A. Gelman, Vol. 29, 119–148. Cambridge: Cambridge University Press.

Ling, Y., H. T. Nefs, W.-P. Brinkman, C. Qu, and I. Heynderickx. 2013. "The Effect of Perspective on Presence and Space Perception." *PLoS One* 8:e78513. https://doi.org/10.1371/journal.pone.0078513.

Losavio, F., L. Chirinos, and M. A. Pérez 2001. "Quality Models to Design Software Architectures." In *Proceedings Technology of Object-Oriented Languages and Systems. TOOLS 38*, 123–135. IEEE.

Maicher, K., D. Danforth, A. Price, L. Zimmerman, B. Wilcox, B. Liston, H. Cronau. 2017. "Developing a Conversational Virtual Standardized Patient to Enable Students to Practice History-Taking Skills." *Simulation in Healthcare* 12 (2): 124–131. https://doi.org/10.1097/SIH.0000000000000195.

Marsella, S. C., and J. Gratch. 2009. "EMA: A Process Model of Appraisal Dynamics." *Cognitive Systems Research* 10 (1): 70–90. https://doi.org/10.1016/j.cogsys.2008.03.005.

Mascarenhas, S., J. Dias, R. Prada, and A. Paiva. 2010. "A Dimensional Model for Cultural Behavior in Virtual Agents." *Applied Artificial Intelligence* 24 (6): 552–574. https://doi.org/10.1080/08839514.2010.492163.

Mascarenhas, S., M. Guimarães, R. Prada, P. A. Santos, J. Dias, and A. Paiva. 2022. "Fatima Toolkit: Toward An Accessible Tool for the Development of Socio-Emotional Agents." *ACM Transactions on Interactive Intelligent Systems* 12:1–30. https://doi.org/10.1145/3510822.

Mayer, R. E. 2004. "Should there Be a Three-Strikes Rule Against Pure Discovery Learning?" *American Psychologist* 59:14–19. https://doi.org/10.1037/0003-066X.59.1.14.

McShane, M., S. Nirenburg, B. Jarrell, S. Beale, and G. Fantry. 2009. "Maryland Virtual Patient: A Knowledge-Based, Language-Enabled Simulation and Training System." *Bio-Algorithms and Med-Systems* 5 (9): 57–63.

McTear, M., Z. Callejas, and D. Griol. 2016. *The Conversational Interface: Talking to Smart Devices*. Switzerland: Springer International Publishing.

Mell, J., and J. Gratch. 2017. "Grumpy, Pinocchio: Answering Human-Agent Negotiation Questions through Realistic Agent Design." In *Proceedings of the 16th Conference on Autonomous Agents and Multiagent Systems*, 401–409. Sao Paulo: International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS).

Motger, Q., X. Franch, and J. Marco. 2022. "Software-Based Dialogue Systems: Survey, Taxonomy, and Challenges." *ACM Computing Surveys* 55 (5): 1–42. https://doi.org/10.1145/3527450.

Mouawad, M., and B. H. Kleiner. 1996. "New Developments in Customer Service Training." *Managing Service Quality: An International Journal* 6 (2): 49–56. https://doi.org/10.1108/09604529610109774.

Muller, T. J., A. Heuvelink, K. van den Bosch, and I. Swartjes. 2012. "Glengarry Glen Ross: Using BDI for Sales Game Dialogues." In *Proceedings AIIDE 2012: 8th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, 167–172. AAAI Press.

Murali, P., F. Nouraei, M. Fallah, A. Kearns, K. Rebello, T. O'Leary, R. Perkins, et al. 2022. "Training Lay Counselors with Virtual Agents to Promote Vaccination." In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*, 1–8. Faro: Association for Computing Machinery, Inc.

Murali, P., H. Trinh, L. Ring, and T. Bickmore. 2021. "A Friendly Face in the Crowd: Reducing Public Speaking Anxiety with an Emotional Support Agent in the Audience." In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*, *Virtual Event, Japan,* 156–163. Association for Computing Machinery, Inc.

Mursion. n.d. https://www.mursion.com/.

Nadeem, M., A. Bethke, and S. Reddy. 2020. "Stereoset: Measuring stereotypical Bias in Pretrained Language Models." arXiv preprint arXiv:2004.09456.

Nakash, T., T. Haller, M. Shekel, D. Pollak, M. Lewenchuse, A. B. Klomek, and D. Friedman. 2022. "Increasing Resilience and Preventing Suicide: Training and Interventions with a Distressed Virtual Human in Virtual Reality." In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*, 1–8. Faro: Association for Computing Machinery, Inc.

Neerincx, M. A., J. v. d. Waa, F. Kaptein, and J. v. Diggelen. 2018. "Using Perceptual and Cognitive Explanations for Enhanced Human-Agent Team Performance." In *International Conference on Engineering Psychology and Cognitive Ergonomics*, 204–214. Springer.

Niemiec, R. P., M. F. Sikorski, and H. J. Walberg. 1989. "Comparing the Cost-Effectiveness of Tutoring and Computer-Based Instruction." *Journal of Educational Computing Research* 5 (4): 395–407. https://doi.org/10.2190/3RH7-3T7A-4JXU-H6CA.

Nwana, H. S. 1990. "Intelligent Tutoring Systems: An Overview." *Artificial Intelligence Review* 4:251–277. https://doi.org/10.1007/BF00168958.

O'Brien, K. H. M., S. Fuxman, L. Humm, N. Tirone, W. J. Pires, A. Cole, and J. G. Grumet. 2019. "Suicide Risk Assessment Training Using An Online Virtual Patient Simulation." *Mhealth* 5:51. https://doi.org/10.21037/mhealth.

Ochs, M., D. Mestre, G. De Montcheuil, J.-M. Pergandi, J. Saubesty, E. Lombardo, D. Francon, and P. Blache. 2019. "Training Doctors' Social Skills to Break Bad News: Evaluation of the Impact of Virtual Environment Displays on the Sense of Presence." *Journal on Multimodal User Interfaces* 13 (1): 41–51. https://doi.org/10.1007/s12193-018-0289-8.

Oh, P.-J., K. D. Jeon, and M. S. Koh. 2015. "The Effects of Simulation-Based Learning Using Standardized Patients in Nursing Students: A Meta-Analysis." *Nurse Education Today* 35:e6–e15. https://doi.org/10.1016/j.nedt.2015.01.019.

Oijen, J. V., W. V. Doesburg, and F. Dignum. 2010. "Goal-Based Communication using BDI Agents as Virtual Humans in Training: An Ontology Driven Dialogue System." In *International Workshop on Agents for Games and Simulations*, 38–52. Springer.

Ojha, S., J. Vitale, and M.-A. Williams. 2021. "Computational Emotion Models: A Thematic Review." *International Journal of Social Robotics* 13 (6): 1253–1279. https://doi.org/10.1007/s12369-020-00713-1.

Othlinghaus-Wulhorst, J., A. Mainz, and H. U. Hoppe. 2019. "Training Customer Complaint Management in a Virtual Role-Playing Game: A User Study." In *Transforming Learning with Meaningful Technologies: 14th European Conference on Technology Enhanced Learning, EC-TEL 2019*, Delft, The Netherlands, September 16–19, 2019, 436–449. Springer.

Peddle, M., L. Mckenna, M. Bearman, and D. Nestel. 2019. "Development of Non-Technical Skills Through Virtual Patients for Undergraduate Nursing Students: An Exploratory Study." *Nurse Education Today* 73:94–101. https://doi.org/10.1016/j.nedt.2018.11.008.

Peng, S., S. Zhu, M. Yang, H. Huang, D. Liu, Z. Wen, X. Li, and B. Fan. 2022. "Adacoach: A Virtual Coach for Training Customer Service Agents." arXiv preprint arXiv:2204.12935.

Pérez, J., E. Cerezo, F. J. Serón, and L.-F. Rodríguez. 2016. "A Cognitive-Affective Architecture for ECAs." *Biologically Inspired Cognitive Architectures* 18:33–40. https://doi.org/10.1016/j.bica.2016.10.002.

Poggi, I., C. Pelachaud, F. de Rosis, V. Carofiglio, and B. De Carolis. 2005. "Greta. A Believable Embodied Conversational Agent." In *Multimodal Intelligent Information Presentation*, 3–25. Dordrecht: Springer.

Prajod, P., M. Al Owayyed, T. Rietveld, J.-J. van der Steeg, and J. Broekens. 2019. "The Effect of Virtual Agent Warmth on Human-Agent Negotiation." In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 71–76. Montreal: International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS).

Rivera-Gutierrez, D. J., R. Kopper, A. Kleinsmith, J. Cendan, G. Finney, and B. Lok. 2014. "Exploring Gender Biases with Virtual Patients for High Stakes Interpersonal Skills Training." In *International Conference on Intelligent Virtual Agents*, 385–396. Springer.

Rossen, B., and B. Lok. 2012. "A Crowdsourcing Method to Develop Virtual Human Conversational Agents." *International Journal of Human-Computer Studies* 70 (4): 301–319. https://doi.org/10.1016/j.ijhcs.2011.11.004.

Sailer, M., and L. Homner. 2020. "The Gamification of Learning: A Meta-Analysis." *Educational Psychology Review* 32 (1): 77–112. https://doi.org/10.1007/s10648-019-09498-w.

Schoenthaler, A., G. Albright, J. Hibbard, and R. Goldman. 2017. "Simulated Conversations with Virtual Humans to Improve Patient-Provider Communication and Reduce Unnecessary Prescriptions for Antibiotics: A Repeated Measure Pilot Study." *JMIR Medical Education* 3 (1): e7. https://doi.org/10.2196/mededu.6305.

Schunk, D. H. 2012. *Learning Theories An Educational Perspective*. Pearson Education, Inc.

Shah, H., B. Rossen, B. Lok, D. Londino, S. D. Lind, and A. Foster. 2012. "Interactive Virtual-Patient Scenarios: An Evolving Tool in Psychiatric Education." *Academic Psychiatry* 36 (4): 146–150. https://doi.org/10.1192/pb.bp.111.035576.

Siemer, J., and M. C. Angelides. 1998. "A Comprehensive Method for the Evaluation of Complete Intelligent Tutoring Systems." *Decision Support Systems* 22 (1): 85–102. https://doi.org/10.1016/S0167-9236(97)00033-X.

SIMmersion. n.d. https://www.simmersion.com/.

Sottilare, R. A., K. W. Brawner, A. M. Sinatra, and J. H. Johnston. 2017. " An Updated Concept for a Generalized Intelligent Framework for Tutoring (GIFT)." GIFTtutoring.org, 1–19.

St-Hilaire, F., D. D. Vu, A. Frau, N. Burns, F. Faraji, J. Potochny, S. Robert, et al. 2022. "A New Era: Intelligent Tutoring Systems Will Transform Online Learning for Millions." arXiv preprint arXiv:2203.03724.

Sveinbjörnsdóttir, B., S. H. Jóhannsson, J. Oddsdóttir, T. Þ. Sigurðardóttir, G. I. Valdimarsson, and H. H. Vilhjálmsson. 2019. "Virtual Discrete Trial Training for Teacher Trainees." *Journal on Multimodal User Interfaces* 13 (1): 31–40. https://doi.org/10.1007/s12193-018-0288-9.

Swartout, W., J. Gratch, R. W. Hill, E. Hovy, S. Marsella, J. Rickel, and D. Traum. 2006. "Toward Virtual Humans." *AI Magazine* 27:96–96.

Swartout, W., B. D. Nye, A. Hartholt, A. Reilly, A. C. Graesser, K. VanLehn, J. Wetzel, et al. 2016. "Designing a Personal Assistant for Life-Long Learning (PAL3)." In *29th International Florida Artificial Intelligence Research Society Conference, FLAIRS 2016*, 491–496. AAAI Press.

Szilas, N., L. Chauveau, K. Andkjaer, A. L. Luiu, M. Bétrancourt, and F. Ehrler. 2019. "Virtual Patient Interaction via Communicative Acts." In *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 91–93. Paris: Association for Computing Machinery, Inc.

Taatgen, N., and J. R. Anderson. 2010. "The Past, Present, and Future of Cognitive Architectures." *Topics in Cognitive Science* 2 (4): 693–704. https://doi.org/10.1111/tops.2010.2.issue-4.

Tanaka, H., S. Sakriani, G. Neubig, T. Toda, H. Negoro, H. Iwasaka, and S. Nakamura. 2016. "Teaching Social Communication Skills Through Human-Agent Interaction." *ACM Transactions on Interactive Intelligent Systems* 6:1–26. https://doi.org/10.1145/2937757.

Titov, N., G. Andrews, I. Choi, G. Schwencke, and A. Mahoney. 2008. "Shyness 3: Randomized Controlled Trial of Guided Versus Unguided Internet-Based CBT for Social Phobia." *Australian & New Zealand Journal of Psychiatry* 42 (12): 1030–1040. https://doi.org/10.1080/00048670802512107.

Topu, F. B., and Y. Goktas. 2019. "The Effects of Guided-Unguided Learning in 3D Virtual Environment on Students' Engagement and Achievement." *Computers in Human Behavior* 92:1–10. https://doi.org/10.1016/j.chb.2018.10.022.

Turk, M. 2014. "Multimodal Interaction: A Review." *Pattern Recognition Letters* 36:189–195. https://doi.org/10.1016/j.patrec.2013.07.003.

Vaassen, F., J. Wauters, F. Van Broeckhoven, M. Van Overveldt, W. Daelemans, and K. Eneman. 2012. "deLearyous: Training

Interpersonal Communication Skills using Unconstrained Text Input." *Proc. of ECGBL*, 505–513.
Van Der Stuyf, R. R. 2002. "Scaffolding As a Teaching Strategy." *Adolescent Learning and Development* 52:5–18.
van der Waa, J., J. van Diggelen, L. Cavalcante Siebert, M. Neerincx, and C. Jonker. 2020. "Allocation of Moral Decision-Making in Human-Agent Teams: A Pattern Approach." In *Engineering Psychology and Cognitive Ergonomics. Cognition and Design: 17th International Conference, EPCE 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Part II*, Copenhagen, Denmark, July 19–24, 2020, 203–220. Springer.
Van der Zwaan, J. M., V. Dignum, and C. M. Jonker. 2012. "A BDI Dialogue Agent for Social Support: Specification and Evaluation Method." In *AAMAS 2012: Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems, Workshop on Emotional and Empathic Agents*, Valencia, Spain, 4–8 June, 2012; authors version. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS).
VanLehn, K. 2006. "The Behavior of Tutoring Systems." *International Journal of Artificial Intelligence in Education* 16:227–265.
VirtualSpeech. n.d. https://virtualspeech.com/.
Washburn, M., D. E. Parrish, and P. S. Bordnick. 2020. "Virtual Patient Simulations for Brief Assessment of Mental Health Disorders in Integrated Care Settings." *Social Work in Mental Health* 18 (2): 121–148. https://doi.org/10.1080/15332985.2017.1336743.
Weidinger, L., J. Mellor, M. Rauh, C. Griffin, J. Uesato, P.-S. Huang, M. Cheng, et al. 2021. "Ethical and Social Risks of Harm from Language Models." arXiv preprint arXiv:2112.04359.
Weston, C., and P. A. Cranton. 1986. "Selecting Instructional Strategies." *The Journal of Higher Education* 57 (3): 259–288. https://doi.org/10.1080/00221546.1986.11778771.
Williams, J. D. 2008. "The Best of Both Worlds: Unifying Conventional Dialog Systems and Pomdps." In *Ninth Annual Conference of the International Speech Communication Association, Brisbane, Australia*.
Yao, H., A. G. de Siqueira, A. Bafna, D. Peterkin, J. Richards, M. L. Rogers, A. Foster, I. Galynker, and B. Lok. 2022. "A Virtual Human Interaction Using Scaffolded Ping-Pong Feedback for Healthcare Learners to Practice Empathy Skills." In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents*, 1–8. Faro: Association for Computing Machinery, Inc.
Yao, H., A. G. de Siqueira, A. Foster, I. Galynker, and B. Lok. 2020. "Toward Automated Evaluation of Empathetic Responses in Virtual Human Interaction Systems for Mental Health Scenarios." In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, Virtual Event, Japan*, 1–8. Association for Computing Machinery, Inc.
Yasavur, U., C. Lisetti, and N. Rishe. 2014. "Let's Talk! Speaking Virtual Counselor Offers You a Brief Intervention." *Journal on Multimodal User Interfaces* 8 (4): 381–398. https://doi.org/10.1007/s12193-014-0169-9.
Ye, P., T. Wang, and F.-Y. Wang. 2018. "A Survey of Cognitive Architectures in the Past 20 Years." *IEEE Transactions on Cybernetics* 48:3280–3290. https://doi.org/10.1109/TCYB.2018.2857704.
Zahabi, M., and A. M. Abdul Razak. 2020. "Adaptive Virtual Reality-Based Training: A Systematic Literature Review and Framework." *Virtual Reality* 24 (4): 725–752. https://doi.org/10.1007/s10055-020-00434-w.
Ziebarth, S., A. Kizina, H. U. Hoppe, and L. Dini. 2014. "A Serious Game for Training Patient-Centered Medical Interviews." In *2014 IEEE 14th International Conference on Advanced Learning Technologies*, 213–217. IEEE.
Zielke, M. A., D. Zakhidov, D. Jacob, and G. Hardee. 2016. "Beyond Fun and Games: Toward an Adaptive and Emergent Learning Platform for Pre-Med Students with the UT Time Portal." In *2016 IEEE International Conference on Serious Games and Applications for Health (SeGAH)*, 1–8. IEEE.
Zlotos, L., A. Power, D. Hill, and P. Chapman. 2016. "A Scenario-Based Virtual Patient Program to Support Substance Misuse Education." *American Journal of Pharmaceutical Education* 80 (3): 48. https://doi.org/10.5688/ajpe80348.

Appendix. Further explanations of aspects in Table 5

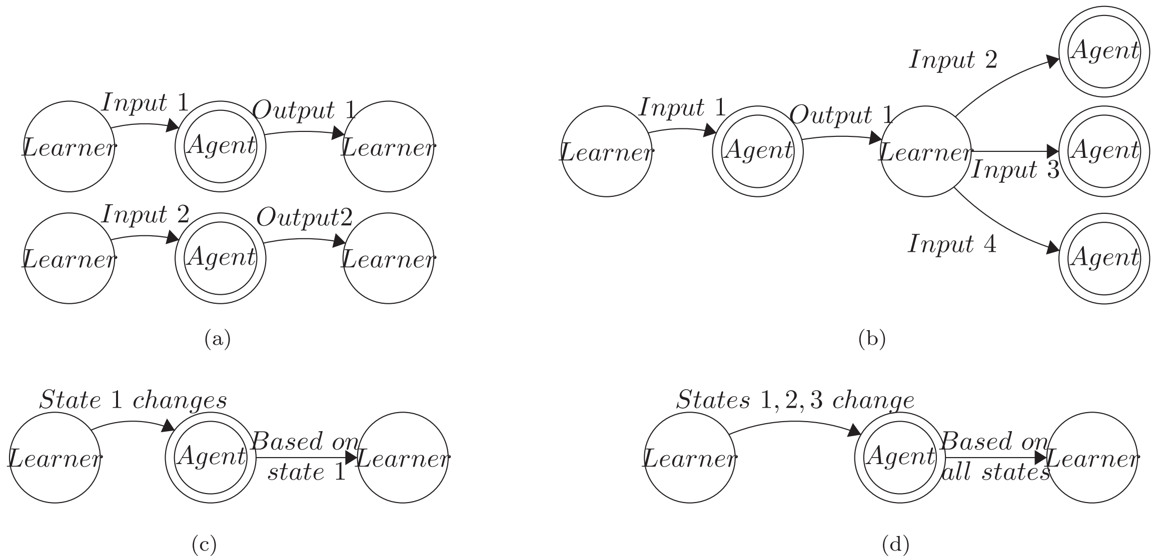


Figure A1. State machine diagrams explaining the four categorization mentioned in Table 5. 'Learner' and 'Agent' refers to their replies. (a) Action decision: The input is directly mapped to output (⚙2). (b) Action decision: a scenario is defined (⚙3). (c) Agent model: limited states changes (⇄2). (d) Agent model: representation of changes (⇄3).