POLITECNICO DI TORINO

Master Degree course in Computer Engineering

Master Degree Thesis

ImprovMate: Multimodal AI Assistant for Improv Actor Training

Supervisor
Prof. Andrea Sanna

Candidate
Riccardo Drago

Academic Year 2024-2025

Abstract

The latest technologies have revolutionized interactions between humans and artificial intelligence, with a significant breakthrough marked by the introduction of ChatGPT by OpenAI in November 2022. New conversational chatbots have been shown to be useful in various fields, not only for simple interactions but also for information search, technical support, and entertainment.

This Master's thesis study investigates the use of AI in creative arts, particularly improvisational theater. This study led to the development of ImprovMate, an innovative system designed to support actors in the creative process and provide a valid alternative to traditional practice.

This system allows users to perform in front of a webcam and practice using AI as an improv partner. A LLM (Large Language Model) can understand and generate coherent text, and adapt to multiple situations. For this reason, ChatGPT has been used to exploit its advanced capabilities, including natural language interpretation and computer vision models. The audiovisual performance of the user is analyzed using these capabilities, which are examined through a motion labeling process on a video dataset. The system adapts ChatGPT to the context of improvisation using carefully constructed prompts. The hallucinations generated by the model are often considered a limitation, but in this case they have been reinterpreted as creative tools to fuel imagination and introduce unexpected elements into improvised scenes.

The design of ImprovMate was guided by a formative study conducted with improvisation actors, to understand their needs, the difficulties encountered in the creative process and the opportunities offered by artificial intelligence in the theatrical context. The insights from this study guided the development of the system, including targeted features, such as dynamic prompts and other support tools, that facilitate experimentation and creative flow.

To evaluate the effectiveness and usability of ImprovMate, a pilot study was conducted to collect preliminary feedback and possible improvements. The results obtained have provided useful indications for future developments of the system, ensuring better adaptation to actors' needs and a more fluid and stimulating interaction.

Contents

Glossary

AI Artificial system that tries to simulate a generic form of intelligence. 1, 7–9, 11–16, 19–21, 23–28, 32–34, 37, 43–46, 48, 54, 57–60, 62, 63, 65, 66

AR Augmented Reality, interactive experience that combines the real world and computer-generated 3D content. 18

ChatGPT AI and machine learning based chat bot developed by Open AI, specialized in conversation with a human user. 1, 14, 25, 26, 49–54

computer vision Set of activities used to extract and understand data from digital images or videos. 1, 7, 24

FPS Frames Per Second. 33

GAI Subset of artificial intelligence that uses generative models to produce text, images, videos, or other forms of data. 12, 13

gesture recognition technology that allows the interpretation and analysis of human movements. 8, 9, 42

hallucinations AI-generated information not based on real data. 1, 11, 27, 32, 34

HTN Hierarchical Task Network, approach to automated planning in which the dependency among actions can be given in the form of hierarchically structured networks. 17

improv Improvisation, form of live theatre where scenes are created spontaneously. 25, 32–34, 36, 37, 39, 42, 44, 57

IoT Concrete objects and places that acquire a digital identity through an internet connection. 18

LLM Large Language Model, a type of AI model used for generating text. 1, 7, 15, 21, 23, 24, 31–34, 49, 53, 55

MediaPipe Open-source framework developed by Google for real-time computer vision and machine learning tasks such as hand tracking, face detection, object detection, and gesture recognition. 42, 43, 49, 50, 52–55

motion labeling Procedure for labeling movements in a video, which is fundamental for analysis and gesture recognition. 1, 49, 51–54

OpenAI Research organization that develops advanced AI models, including ChatGPT. 1, 31–33, 42, 43, 48, 50, 65

similarity Sentence Similarity is the task of determining how similar two texts are. 51–55

storytelling The art and practice of conveying a narrative through structured or spontaneous techniques. 7–9, 11, 16, 18–23, 32–34, 36, 48

TTCT Torrance Tests of Creative Thinking, series of standardized assessments developed by Ellis Paul Torrance to measure an individual's creative potential. 14

Turing Test Test of a machine's ability to exhibit intelligent behaviour equivalent to that of a human. 15

Chapter 1

Introduction

Improvisational theater, at its core, is a storytelling art, an expressive practice in which narratives are spontaneously created through dialogue, movement, and emotion. In this dynamic form of performance, actors are able to showcase their creative abilities, using their intuition and instincts to collaborate in the creation of an engaging, ever-evolving story. This form of storytelling is not scripted; instead, it thrives on the unpredictability of human interaction, where every gesture, pause, and feeling contributes to the unfolding of the story.

In recent years, AI (artificial intelligence) has begun to be increasingly used in fields related to creative disciplines, such as in the processing of narratives. The most advanced AI systems, such as LLM (Large Language Models), are capable of generating high-quality coherent stories and texts. Instead, models that also possess computer vision are able to understand human movements and behaviors starting from images or videos. These features open the doors to infinite possibilities and these capabilities can be integrated into the creative process, not to replace the human contribution but providing a collaborator capable of providing a different valid point of view.

In this art, even the most talented artists can find themselves struggling during performances, overwhelmed by unexpected twists that undermine the coherence of the narrative. Training for these situations is not easy, and each actor has to find partners to train with specific situations and follow all the techniques underlying improvisational performances.

This thesis explores the supporting role of AI as a creative partner in storytelling in the context of improvisational theater. The aim of the work is to develop a system that not only supports but also enables improvisation in the absence of companions. Using the latest technologies, the tool provides narrative cues in real time, tracks key elements of the story, and builds dynamic dialogues. To provide a relevant background to the project, interviews are conducted with experts in the field to properly direct development.

1.1 Scenario

In traditional improvisational theater, actors rely on their instincts, experience, and suggestions from a live audience to guide their performance. However, not everyone has easy access to a troupe of actors or a live audience to practice. This creates a scenario where actors may not have effective training opportunities, which can limit their growth and creativity. The problem addressed in this work is how to provide an accessible, reliable, and interactive training tool that supports improvised storytelling without the need for extensive human resources or specialized equipment.

Furthermore, such a narrative experience is not only important for theater professionals; it also improves creativity, communication, and trains the ability to think in unexpected conditions. The approach is not limited to the specific field of application, but remains open to anyone who wants to challenge themselves in such situations.

The idea of using AI to assist in creative processes is both innovative and practical. AI-based tools have a very different way of reasoning than humans, yet are similar in certain aspects. This difference can be the basis for constructive collaboration, capable of improving the actors' skills and confronting themselves with a different point of view.

Finally, with the advances of recent years, particularly in language understanding and movement recognition, the potential to create a system that interacts with the actors in real time becomes feasible. This approach can make training available to anyone interested in improvisation to practice and improve, regardless of their access to traditional performance spaces or partners.

1.2 Main Contribution

The main contribution of this thesis is the development of ImprovMate, an AI-assisted system designed to support improvisational theater. ImprovMate offers two distinct interfaces:

- **ImprovMate:** a step-by-step interface that gradually builds the story by providing structured narrative cues and tracking key elements such as characters, settings, and objects. Additional exercises are proposed to help train specific skills.
- **ImprovMate RT:** a real-time interface that uses gesture recognition and dynamic feedback. It simulates live performances through immediate interaction with an AI improvisation partner and an AI audience.

These interfaces take a different approach, offering actors flexible training tools that adapt to different creative needs and performance styles.

To achieve this aim, this thesis initially refers to existing research on creativity and its relationship with AI. The various existing storytelling techniques are analyzed, focusing on those that exploit cutting-edge technological devices. To understand the needs of the target audience, a formative study with actors is conducted to gather information related to the field of interest. The results are used to guide the design of the system and make it as user-friendly as possible.

During the development of the system, the two interfaces mentioned above are designed and implemented, using the latest AI models, gesture recognition and interaction technologies. Experiments were conducted to evaluate the system performance and finally feedback was sought from actors to guide future research.

1.3 Thesis Structure

The thesis is structured as follows:

- **Chapter 2 - Related Work:** this chapter reviews previous research in areas including AI to increase creativity, AI in improvisational theater, interactive storytelling, AI-based storytelling and motion-driven storytelling.
- **Chapter 3 - Formative Study:** this chapter presents initial studies conducted to understand the role of AI in improvisation, the relationship between age and willingness to use AI, and the impact of randomness on narrative coherence. The results of the study define the design goals for the system.
- **Chapter 4 - System Overview:** this chapter provides a detailed description of ImprovMate, outlining the architecture of the system and the features of the step-by-step and real-time interfaces.
- **Chapter 5 - Interrogative Study:** the capabilities of the large language model in analyzing movements are examined in this chapter, along with a description of the experimental approach and the results obtained.
- **Chapter 6 - Pilot Study:** this chapter evaluates the system in a real-world setting by analyzing its performance based on feedback collected from actors.
- **Chapter 7 - Conclusion and Future Work:** the final chapter summarizes the key contributions of the thesis and discusses possible directions for future research.

Chapter 2

Related work

The integration of AI in creative domains has received increasing attention in recent years. AI-based systems are being explored as tools to enhance creativity, provide interactive storytelling experiences, and, in more niche cases, support improvisational theater.

This chapter reviews the relevant literature on technology-assisted storytelling techniques, with a particular focus on AI-based tools. Its creative capacity is assessed, considering the most varied approaches.

First, this chapter discusses the proposed considerations with respect to the creative component of AI, which is also used as a support tool to refine artistic expressions and overcome creative blocks. Then, the applications of AI in improvisational theater are explored, a field of art that can harness AI hallucinations for something new. Next, interactive storytelling is considered. This term refers to a domain that uses and integrates various technological approaches to storytelling to create engaging experiences for users. Finally, motion-based storytelling can be considered a part of this last branch. This refers to all the approaches in which the users' movement is captured through various methodologies and interpreted and exploited to carry on the narration of stories.

With the aim of proposing an effective AI-based system to support improvisation actors, past work is analyzed trying to identify the main opportunities and limitations.

2.1 AI to Increase Creativity

Since 1998, several researchers have examined the possible capacity of artificial intelligence to produce creative ideas [1]. Boden's work explores the intersection between human creativity and artificial intelligence, offering insights into how computational systems can both model and improve creative processes. In this publication, different ways of producing novelty are explained.

- **“P-creativity”** (psychological creativity), which refers to ideas that are new to the mind of the individual;
- **“H-creativity”** (historical creativity), which introduces ideas that are new to the whole of history.

As for the different types of creativity, it is possible to distinguish three types:

- **combinatorial creativity**, concerning the new (and unlikely) combination of familiar ideas;
- **exploratory creativity**, generating ideas that are not new, but unexpected, by exploring conceptual spaces;
- **transformative creativity**, modifying conceptual spaces for the generation of new ideas.

Boden examines how AI can simulate creative processes, explaining how combinatorial creativity is emulated by AI systems by mixing existing concepts to create new ones, such as in the creation of jokes. In terms of exploratory and transformative creativity, the exploration and modification of conceptual spaces can be described through computational concepts, as in the creation of new musical genres starting from key traits of other composers.

Boden believes that AI can model and, in some cases, exhibit creative behaviors, but the evaluation remains subjective in any case, as it depends on human judgment of novelty. Finally, the author believes that even if AI can generate novelties that perplex or repel us, it is still far from convincing us that they have value.

With technological advancement, AI has become part of everyday life, and different approaches are proposed with which these tools can help humans in many creative activities.

Wu et al. [40] introduce the concept of “AI Creativity”, highlighting the collaborative potential between humans and artificial intelligence to improve creative processes, exploiting each other's strengths (see Figure 2.1). To exploit this approach, the authors propose a human-AI co-creation model, a circular process model that includes six phases: perception, thinking, expression, collaboration, construction, and testing. This model illustrates how AI introduces new possibilities in each phase of the creative process. When talking about AI creativity, it should be remembered that it depends on the material on which it is trained and, therefore, it is influenced by humans. For this reason, AI must be a tool to lower the entry bar of various fields of study, allowing humans to focus on the creative part, leaving the more complex or time-consuming tasks to AI.

Haase and Hanel's [10] work examines the creative capabilities of Generative Artificial Intelligence (GAI) chatbots, comparing their performance to human creativity. The study evaluates ideas generated by six GAI chatbots, compared to

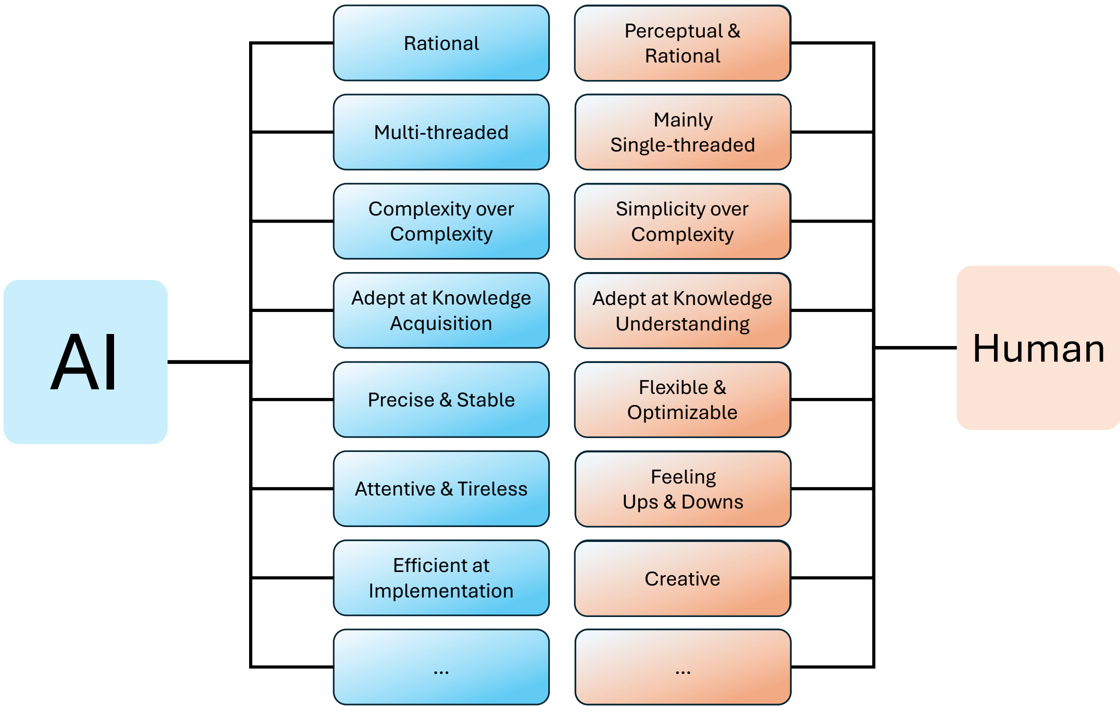

Figure 2.1: Humans and AI are complementary: humans take charge when creativity, strategic thinking, or empathy is required, while AI excels at handling routine or optimization tasks where compassion is less of a factor. Reproduced from [40].

ideas generated by humans. Both human judges and a specially trained AI evaluated the quality and quantity of these ideas.

Despite the differences in idea generation techniques, the results show no qualitative difference between AI and ideas generated by humans. The findings imply that GAIs can be useful as helpers in creative pursuits, possibly boosting human creativity by providing a variety of points of view and ideas.

It is emphasized that GAIs excel for everyday, *small-C* creativity, and are not as effective for *Big-C* achievements, related to contexts of large scope and impact. Despite this, since humans typically lack *Big-C* creativity, which is rare for a few brilliant minds, it can be said that GAIs are creative "as much or as little as humans".

Inie et al. [14] explored how creative professionals perceive the evolution of generative AI and its integration into their workflows. The study reveals that creative professionals have mixed feelings about GAI, recognizing its potential to improve productivity but expressing concerns about its impact on the creative process and job security. Participants reflect on the essence of creativity, questioning how GAI fits into traditional creative paradigms and whether AI-generated content can be considered truly creative. Although it is acknowledged that generative AI is dependent on human creations and input, it is acknowledged that it can be considered creative. AI lacks intention and the application of experience in the creative process, but otherwise the act of combining to create something new is not that different from what humans do, albeit on a bigger scale.

Professionals also express concerns about their future potential employment and the potential for AI to weaken the “creative muscle”, but they recognize that generative AI has the potential to become a fundamental tool.

However, the study by Hubert et al. [12], concludes that current AI language models, particularly *GPT-4*, show a higher creative potential than humans in divergent thinking tasks. Especially for the fluency of responses, AIs are more original and elaborate.

Other studies confirm this trend [9], evaluating the performance of ChatGPT using the Torrance Tests of Creative Thinking (TTCT).

Magni et al. [20] examined whether the identity of a creator, human or AI, influences how people evaluate the creativity of an artifact.

The study concludes that people can be biased towards AI-generated artifacts. These are perceiving as less creative than humans-made ones, partly due to the belief that AI puts less effort into creation.

Other works [39] explore the evolving relationship between AI and human creativity, highlighting the potential for collaboration, but also stressing the need for ethical guidelines in generative AI. The core principles are proposed to guide the responsible and ethical use of AI in creative fields, ensuring that AI serves as a useful tool rather than a disruptive force. And finally, maintaining human oversight is important to ensure that creative outcomes align with human values and cultural contexts.

2.2 AI in Improvised Theater

The latest developments in artificial intelligence have allowed us to provide extremely fast and effective tools in text generation. By introducing these in the creative domain, many researches have tried to exploit the capabilities of these models in the theatrical field. Various publications have addressed the writing of plays through the use of AI [27,28,33,34], however, a more niche approach concerns the integration of these tools in improvised theater. In these papers, it is examined how AI can work together with humans, both trying to use these tools to improve the creative process but also as active participants in live performances.

One of the most active and interested researchers in these specific applications is P. Mirowski, who has started exploring the possibility of integrating artificial intelligence in live theatrical performances [21] (see Figure 2.2a). The authors’ goal is to work in a way that enables a future where AI can collaborate with humans.

One of his first studies examines two different AI-based performers, designed to collaborate with actors and perform improvised theater, without any script. These agents were designed using machine learning techniques to generate responses that were appropriate for the context and were tested in front of an audience to evaluate their impact. This approach is innovative, but at the same time limited by the capabilities of the AIs, which are unable to correctly interpret complex dialogues and interactions such as those of an improvisation performance. The technology used is not yet sufficiently advanced in speech-to-text capabilities and is not able to integrate nonverbal cues used by actors on stage, showing difficulties in maintaining narrative coherence during the performance.

(a)

(b)

Figure 2.2: (a) A human actor performs on stage with A.L.Ex., personified as a robot. The user interface, which displays the results of the speech recognition and the response generated by the machine, is projected behind the performers. Reproduced from [21]. (b) Improbotics cast performing AI-powered improvisational theater. An actor wears a headset connected to a radio system that receives text-to-speech from lines generated by LLM (and edited by an operator). The result of the speech recognition is used as a prompt for the LLM. Reproduced from [4].

Subsequently, the same authors propose this idea again [22], examining the use of AI, specifically deep learning models trained on extensive datasets, to participate in theatrical performances alongside actors. In this approach, actors perform AI-generated lines received via earphones, and the audience finds itself involved in a Turing Test, tested in recognizing the lines generated by AI or not. This collaboration aims to improve the creativity of actors who find themselves having to manage unexpected situations, introducing new dynamics. Anonymous feedback from the audience is also collected, highlighting how problems regarding the consistency of the narrative and the timing of the sentences generated by AI are still unsolved. To make the performance more realistic, a human operator is needed to manage the generated dialogues and provide them to the actors. For further feedback on this matter, the authors themselves also held workshops [26], involving interested actors and staging additional live performances.

The subsequent work sees AI used to help actors in multilingual performances [25], to increase inclusion with an innovative approach. The authors use a *GPT-2* neural network, speech recognition, text generation, and text-to-speech technologies, but these are not enough to avoid errors of various types. The models used are trained on datasets of English texts, and for this reason they cannot effectively manage other languages. Although translation errors introduce a comical edge to the performance, they can also lead to confusion or unwanted offense, given the lack of sufficient context.

With the introduction of *GPT-3*, Mirowski's experiment is being re-staged [3], using an AI that can keep track of the plot and characters used. Audience feedback has shown that an AI narrator is preferred to having the AI on stage. The actors have appreciated the randomness introduced by the AI, but there are still obvious problems with story development. The system used is not capable of consistently developing the personalities of the characters and is not capable of providing valid leads for the initial construction of the story, leading the actors to prefer AI-free approaches.

In one of the latest studies [4], the authors try to expand the use of AI, interacting with more than one actor at a time (see Figure 2.2b). It is noted that, using *GPT-3* with sufficient prompting and context, it is possible to allow AI to consider more participants on stage, but this is not sufficient. The audience remains dissatisfied with the use of AI as a creative storyteller, highlighting how it is often unable to carry on the narrative in a coherent way. A system based only on speech recognition is not able to capture all the contextual details given by the actors and for this reason remains incomplete. The authors postpone to future works the search for ways to introduce sufficient context without causing excessive delay that would damage the quality of the performance.

2.3 Interactive Storytelling

Although storytellers have been making and telling stories since the ages of time, interactive storytelling is a newer approach. This approach promotes tools and resources to help users create stories. In recent years, many technological improvements have further advanced the field.

Several papers examine the construction of frameworks capable of providing an interactive experience to the user for the creation of stories for their own enjoyment. Bostan's paper [2] highlights the importance of communication processes in conjunction with technological tools. The authors argue that player enjoyment can be maximized by personalizing the gaming experience in real time, adapting both the form and content to the player's individual preferences.

The landscape of storytelling applications has seen several innovative contributions through the use of different types of technologies, such as those that integrate physical and virtual elements.

One of the first approaches is that of Grasbon et al. [8] in which V. Propp's morphological analysis of fairy tales is examined in interactive narratives. After identifying the main narrative elements, the authors developed a narrative engine designed to combine them dynamically. This system is designed for a mixed reality approach and allows the user to interact with the modules used for the story construction. Each interaction is mapped to a specific event and the user can choose between different proposed options. Despite the interesting approach, it is not a story generation system, but rather an interactive story creation system whose story is generated from the scenes that must be manually inserted.

Cavazza et al. [5] propose an interactive narration based on characters, where each character has an initial plan and the combination of the plans of each character leads to the development of a story. Without user input, the story continues toward one of the possible endings generated by the intertwining of these variables. When the user interacts, they can influence in real time the outcome of the characters' actions, triggering the replanning of the plot and therefore a different outcome. This approach uses the Unreal Tournament game engine and Hierarchical Task Network (HTN) scheduling to model autonomous character behaviors within an interactive narrative. Real-time adaptability provides an immersive experience, but the computational intensity of HTN scheduling poses challenges in maintaining system performance, allowing only simple stories with a maximum duration of 3 minutes.

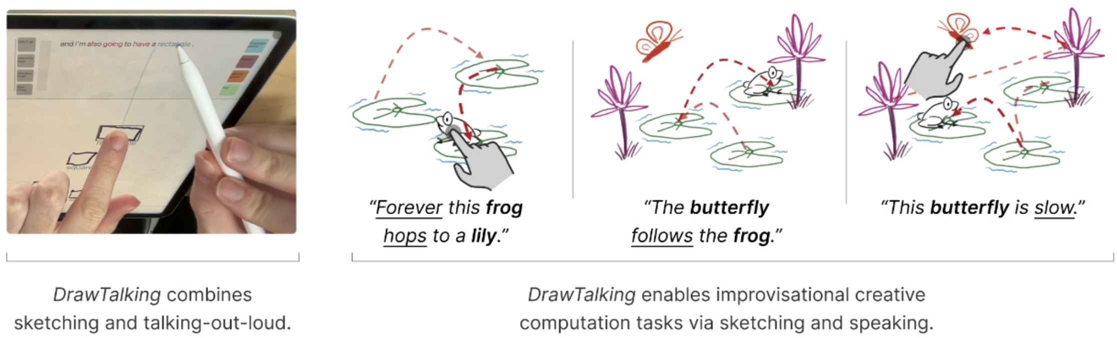

Figure 2.3: Users narrate a story that brings their imagined world to life, with the system integrating this narrative with the designs already in place. Reproduced from [35].

Rosenberg's *DrawTalking* [35] introduces the ability to create and control interactive worlds through a combination of sketches and verbal narration (see Figure 2.3). This approach allows users to tell a story for the creation of a world they imagine, and the system connects the story to what has been designed so far. Each element is labeled, and it is possible to interact with it through a language processing system that extracts information from speech. The proposed prototype has limitations, with respect to the recognized vocabulary and the recognition system, which has room for improvement.

Interactive storytelling is essential to ensure entertaining experiences even for games or similar systems. For example, a storytelling system like *PaSSAGE* (Player-Specific Stories via Automatically Generated Events) [38] adapts narratives to the individual preferences of the player by learning their preferred play style. This approach improves player engagement by dynamically selecting customized story events according to individual styles, demonstrating the value of personalized storytelling in role-playing games.

The various past researches focus on different types of audiences, including young people, trying to exploit the creative spirit of children.

StoryMakAR [7] provides an accessible plug-and-play system that combines electromechanical devices with virtual characters to create stories. It focuses on the use of low-fidelity (lo-fi) materials to improve accessibility, instead of using expensive and traditional toolkits. The combination of these technologies also introduces challenges, such as users without prior knowledge being unable to fully utilize the potential. Despite this, users have been enthusiastic about it and have even requested the addition of more tools such as voice recognition to enrich the stories they create.

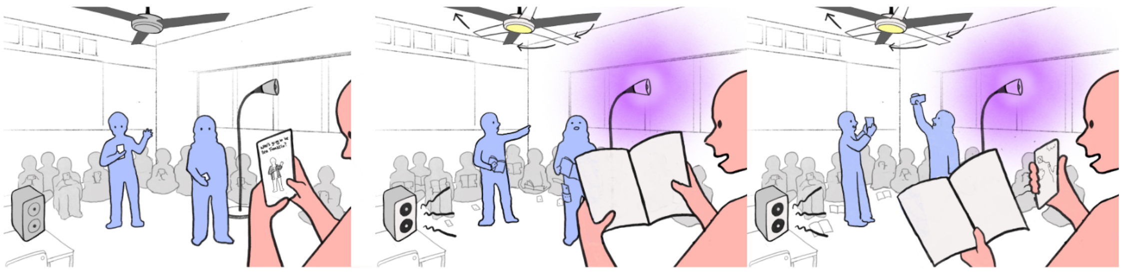

Figure 2.4: Jigsaw merges augmented reality with IoT devices to create a truly immersive experience. For example, users can choose their character simply by waving their hand; as the story unfolds through narration, certain keywords trigger changes in the physical setting, visible through devices like smart lights, fans, and speakers; and virtual elements such as kites, clouds, and sparkles appear within the augmented reality view. Reproduced from [43].

Another approach is taken in *Jigsaw* [43], which uniquely combines mobile augmented reality (AR) with readily available Internet of Things (IoT) devices, creating an immersive storytelling experience (see Figure 2.4). The main issue with using *Jigsaw* is sensory overload, which can make it difficult for users to keep up with the rapid changes in the story. Unlike *StoryMakAR*, this approach is certainly more accessible, as it allows users to use the system without specific devices or advanced programming skills.

However, the configuration of *Jigsaw* needs expert designers and engineers. For this reason and because of the need for expensive equipment, accessibility is limited. Additionally, creating complex or long stories is challenging due to limited support for duplicating triggers, scenes, or behaviors. This makes *Jigsaw* more suited to creating short stories, rather than creating extended narratives.

2.4 AI-based Storytelling

With technological advances in recent years, applications have increasingly been equipped with AI-based capabilities. Artificial intelligence-based storytelling systems have significantly advanced the way narratives are created and experienced.

The research by He et al. [11] led to the creation of a novel framework that takes advantage of existing video clips to generate coherent narrative videos (see Figure 2.5). The system generates video aligned with the storyline requested by the user. This process is guided by the retrieved motion structures and text prompts. In addition, the user can specify the desired character identities via text prompts, ensuring consistent character representation throughout the video. The innovation inherent in this framework is limited by the quality and diversity of the generated videos, which are influenced by the video clips available in the retrieval system.

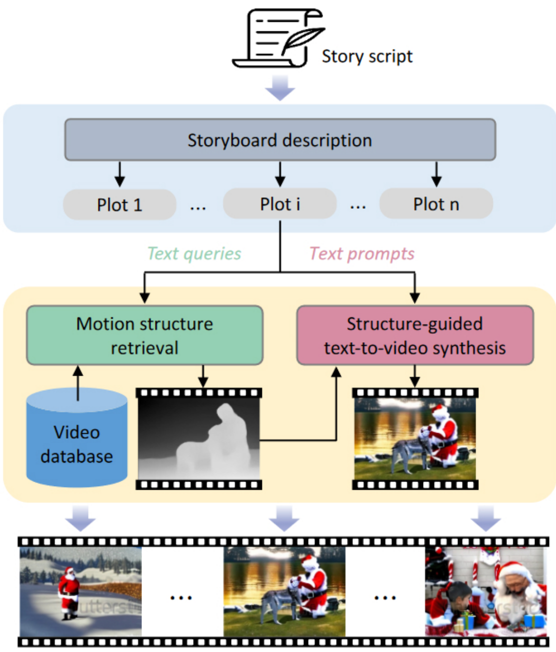

Figure 2.5: Starting from a text script of a story, the system first identifies the main plot points and converts their descriptions into text queries and prompts. Then, each plot is transformed into a video clip generated using two modules: a video retrieval system and a structure-driven text-to-video model. Reproduced from [11].

Other works such as *Rolling the Dice* [36] integrate generative AI to improve storytelling in tabletop role-playing games such as Dungeons & Dragons (D&D). The authors introduce generative AI as a companion to storytelling, helping to create narratives and character development, thus enhancing the experience provided. The paper recommends design guidelines for creating tools that use generative AI in interactive storytelling, raising questions about the potential impact on player immersion and cognitive load.

The *SAGA* system [37], on the other hand, uses AI to facilitate collaborative writing between multiple people, in an asynchronous manner. The approach involves a turn-based system, where each user can contribute their part of the story after deciding on an initial prompt of common agreement. Although the asynchronous approach can be useful in many cases, it is not intuitive, as it must be managed entirely by the users. They also need more help in creating effective prompts for this purpose, and must always keep the story in mind to remain coherent.

Storybuddy [44] introduces a system for children that allows parents to be involved (see Figure 2.6). They can customize the settings as they wish to ensure the best possible experience, keeping track of their progress in educational fields. Children can interact with the conversational AI agent through a speech recognition tool, which is not exactly accurate. The user study administered shows how there is room for improvement, to improve the quality of the content generated by AI and to ensure more accurate customizations based on the desired objectives.

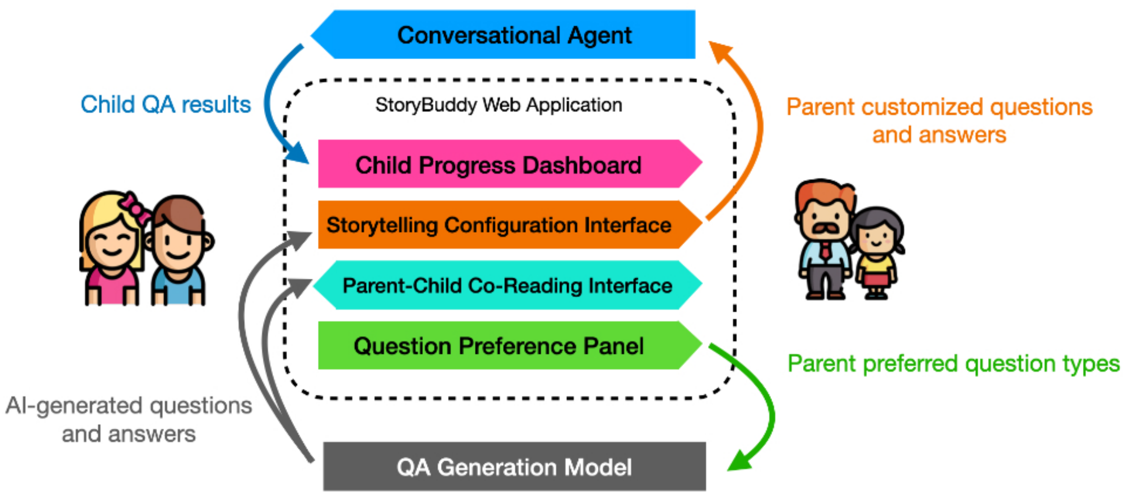

Figure 2.6: A parent selects and customizes a story, then, during the co-reading session, the AI module dynamically prompts questions, captures and evaluates the child's responses, and logs progress on a dashboard. Arrows indicate interactions directed to parents (right) versus children (left). Reproduced from [44].

Support for children is also provided by *StoryDrawer* [42], trying to help them during the narration of their stories. This system proposes suggestive ideas and transforms everything that is mentioned in the narration into drawings, helping them to continue even if they do not know how to draw specific objects. The results show how children see AI as a partner capable of overcoming the anxiety of creative block, overcoming the difficulties of storytelling. In the future, the authors reserve the possibility of testing whether this approach is able to develop children's creativity in the long term.

Similarly, *Wordcraft* [41] proposes the use of an AI as a writing partner in a context not only aimed at children. Using an LLM, the user constructs a story in cooperation with a partner able to increase creativity. The study highlights great opportunities, but AI remains limited in understanding complex narrative structures and is inconsistent with longer stories.

With a purely educational approach, *Storyfier* [32] proposes users to read stories generated by the AI to learn specific words. Subsequently, to consolidate the new notions, the AI helps the user to generate stories that include the target words. Despite the proactive approach, the user study does not show a significant improvement in learning and the system requires further improvements, improving the quality of the stories by avoiding ambiguities in the use of words.

2.5 Motion Guided Storytelling

There are also systems that can exploit movements, gestures, or body actions to create stories, allowing a level of engagement that is unattainable with other approaches. Through technologies developed in the last decades, this type of storytelling has been made possible with various innovative methods.

With a very basic approach, *Shape Your Body* [16] offers users a framework capable of processing body movements in real time, using motion sensing technology. Users can control virtual puppets through movements of their entire body, but facial expression recognition or interaction with multiple users is not supported. Despite the imperfect accuracy, this idea lays the foundation for interesting developments.

In this field, many researchers choose children as the target audience for their work, as some projects mentioned above.

Lin et al. [19] present an approach capable of involving and stimulating the most creative component of children. Through this project, children are able to bring their drawings to life, moving them through their fingers. Interaction is made possible through a tablet or alternatively, children can define movements and deformations by clicking directly on the images that have been projected. Movement is detected through infrared light and a camera capable of detecting contact with the screen. It is possible to note how this approach has encouraged collaboration between several children, testing their narrative skills, and creating entertaining stories with the characters they have created. Unfortunately, the system remains dependent on the specific technologies and configurations used and is not able to involve movements other than those of the fingers.

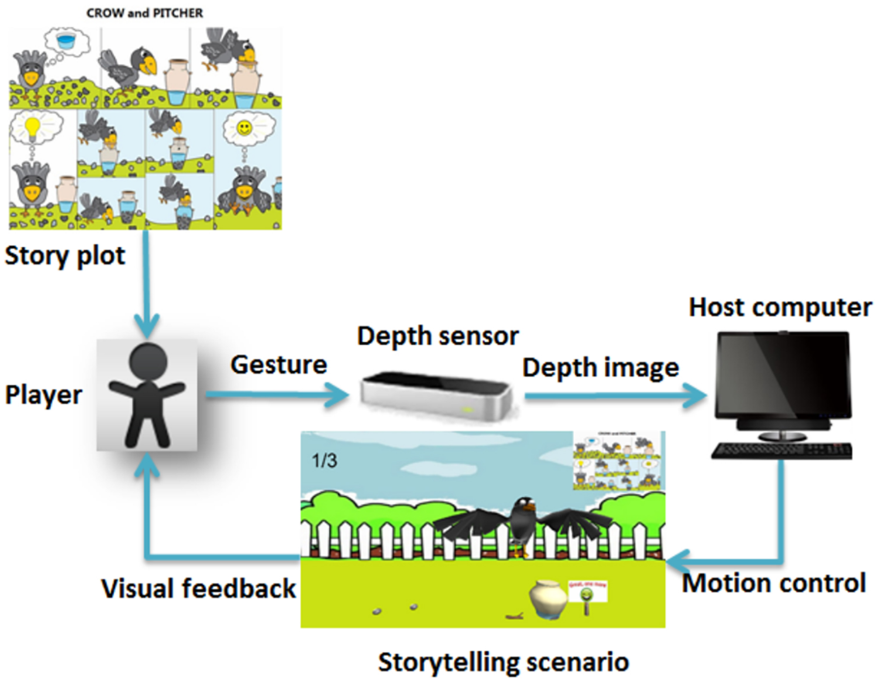

Figure 2.7: The plot of the story is provided as an initial hint to the user. Then, players use hand motion to manipulate the avatar via a depth motion sensor device, which can track and recognize hand gestures. Finally, as a feedback to the user, the avatar's response animation is provided to the players so that players can adjust their hand motion/gestures to continue with the narrative. Reproduced from [17].

Puppet Narrator [17], similarly, makes story narration possible and allows the animation of a virtual avatar through hand gestures (see Figure 2.7). Technologies capable of tracking hand movements are used, which makes the system dependent on the sensors used. Through the latter, children can use their gestures to guide the character in the story, but the approach remains poorly customizable. The range of movements that users can use to direct the storytelling process is limited by the fact that these types of system can only identify hand gestures that are part of a predefined set. This limitation reduces the adaptability and originality that people can add to their stories. Furthermore, this system only provides one story, which reduces the repeatability and diversity of the storytelling experience.

Ready...Action! [6] extends the approach by allowing children to act out scenes using their own body movements. The system captures physical movements via an external system using markers and the *OptiTrack motion capture suite* [31], which mirrors these actions in real time via a graphically supported cartoon character. This approach provides a highly interactive and engaging storytelling experience.

A limitation of *Ready...Action!* is that it only allows one child to act at a time and requires another child to control the keyboard to manage the recording of the animation. Initial user studies showed that children required more than the 15 minutes allotted for the system tutorials, not allowing for immediate use. Another significant limitation is the reliance on external equipment, such as motion capture systems and physical markers, which can be expensive and cumbersome, limiting the accessibility and scalability of the system.

To provide a satisfactory set of gestures, Kirstler et al. [15] conduct a user study asking participants to propose intuitive gestures to trigger specific actions in the proposed stories. This study highlights how variable the possible proposals of users are, and for this reason all systems that want to rely on this mechanics must be flexible and accommodate different preferences.

2.6 Key Challenges and Opportunities

Although AI approaches offer great opportunities and challenges, the works cited in this chapter show that there is still room for development. LLMs have expanded the potential of AI as a tool for creation and support, as well as in the analysis of movements. However, there are still many limitations that highlight the subtle difference between humans and AI. These current systems are unable to effectively emulate the spontaneity and coherence that humans integrate into their narratives. And another crucial point is the possibility of having fluid interactions with little waiting time to be able to integrate these tools into practices such as live improvisation.

Furthermore, while existing AI models can generate engaging cues and suggestions, they lack a deep understanding of artistic intent. Many essential aspects of improvised performance, such as timing and spontaneity, are complex concepts for AI. Current models struggle to replicate the diverse rhythms and dynamic interactions that naturally occur between human performers. Additionally, the lack of shared physical and emotional context makes it difficult for AI to contribute meaningfully in real-time performance contexts.

One of the persistent challenges in AI-driven storytelling is narrative coherence, ensuring that dynamically generated stories maintain a logical progression while allowing for creative freedom. AI-generated content risks losing continuity, especially in long-form narratives.

To address these issues, human intervention often provides satisfying experiences, but this does not create self-sufficient tools.

A hybrid approach that combines user input and AI-generated content could provide an interactive and customizable approach for any eventuality. Trying to shift the focus from replacing actors to creating an AI-based collaborator could provide new approaches that are valid for complex narratives.

Systems that enable narration by incorporating extra inputs, such as gestures, often rely on dedicated hardware. Motion capture devices are often used as tracking markers for motion capture, which can limit their usability for the audience. Additionally, using a fixed set of gestures does not consider the rich diversity of movements that can be found in different theatrical contexts.

Recent innovations in marker-free motion analysis using computer vision and LLM-based visual understanding offer new possibilities. Artificial intelligence can now interpret body movements in real time using only a webcam, making AI systems more accessible and scalable for a wide range of artists.

Chapter 3

Formative Study

To inform the design of ImprovMate, a formative study was conducted with 15 performers from an improv club, ensuring a diverse range of perspectives on the integration of AI into improvisation. The participants varied in gender (F: 5, M: 10), age (18-25 y/o: 6, 25-35 y/o: 5, 35+ y/o: 4), experience (11 beginners, 2 experts) and years of practice (less than 1 year: 3, 1-3 years: 10, 3-5 years: 1, more than 5 years: 1), allowing a comprehensive understanding of different needs and expectations.

Data were collected through 13 questionnaires and two in-person interviews, which provided insight into how AI could support their creative process, as well as potential concerns or limitations they perceived. The key themes of the data are summarized below.

3.1 Correlation between Age and Willingness to Use AI for Practice

Despite the limited number of participants, an interesting correlation can be observed between the age of the participants and their willingness to the use of AI for improvisation practice (Figure 3.1).

Younger generations, including participants aged 18 to 25 years, were generally more enthusiastic about incorporating AI tools into their exercise routines, and all rated the proposed idea positively. This is probably due to their familiarity with technology and the use of chatbots such as *ChatGPT*, as they also mentioned in their responses.

In contrast, the scores are lower for older participants (age groups 25-35 and 35+ years) who demonstrated greater skepticism toward the use of AI in their training. Their concerns often stemmed from a preference for traditional methods, apprehension about the ability of AI to replicate human-like interactions, or concerns about the potential impact of AI on the naturalness and spontaneity of performance. Although some older participants recognized the potential of AI for specific tasks, such as providing feedback or tracking narrative elements, they were less likely to consider it a critical component of their training process.

While this correlation suggests a generational influence on the acceptance of AI in creative practices, the current results are based on a limited sample of participants and may not represent the broader trend. A larger study with more diverse participants is needed to confirm this pattern and explore other possible factors that influence openness to adopt AI.

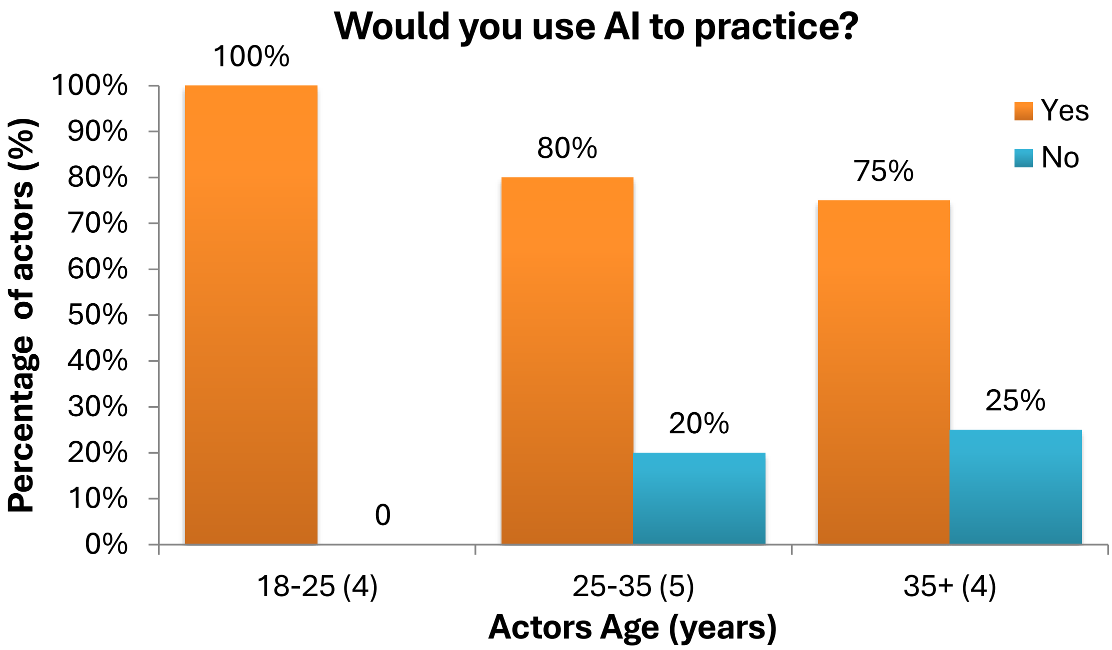

Figure 3.1: Correlation between age and actors' openness to using AI for practice. This chart shows the percentage of actors in three age groups (18-25, 25-35, and 35+) who were either willing or unwilling to integrate AI tools into their practice routines.

3.2 Role of AI in Improvisation

The study revealed different perspectives on the role of AI in improvisation training, with references to commonly used tools such as *ChatGPT* (Figure 3.2).

Many performers have recognized the potential of AI to provide *immediate feedback* and *personalized training*. This technology can be a useful support, especially when human training partners are not available. In fact, some participants have referred to their experience with modern chatbots and have perceived AI as a valid alternative in solo practice. However, the results clearly show that the presence of

AI as an actor on stage is not considered as other options. This consideration could be traced back to the still widespread mistrust towards new technologies, not able to replicate the experience and adaptability of human performers.

In other cases, the participants thought that it may be possible to simulate the input of the audience. AI could provide ideas as the story unfolds, just as the audience does during a performance. However, concerns arose about excessive reliance on AI, which some feared would detract from the natural spontaneity of human-led performances.

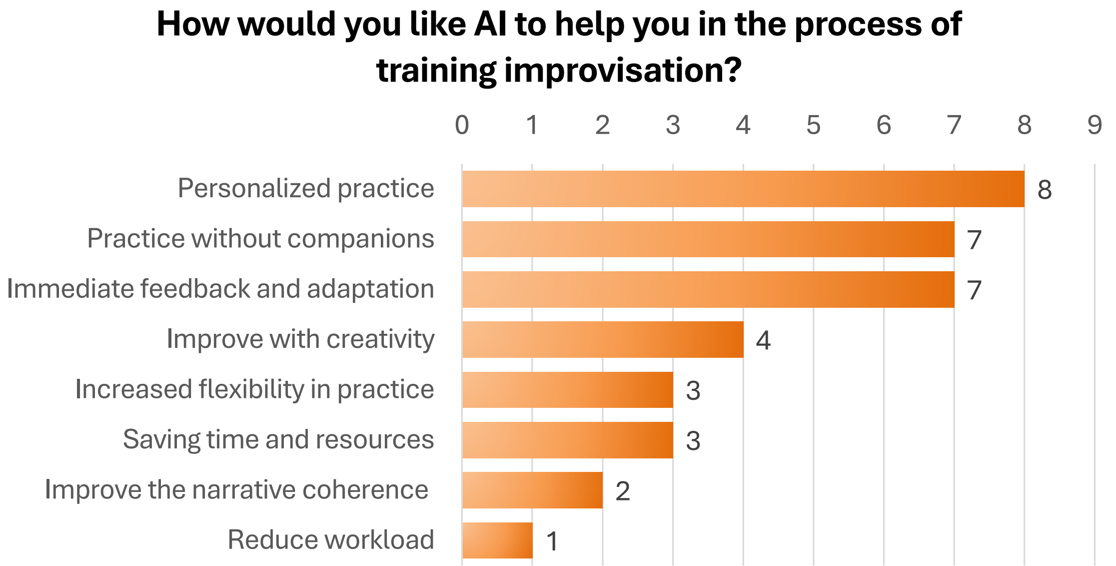

Figure 3.2: Actors' preferred types of assistance from AI, with the option to select multiple responses. A total of 13 actors participated.

3.3 Randomness and Narrative Coherence

AI hallucinations and randomness can often present challenges, but not necessarily in the creative domain. Many participants noted that unexpected stimuli can be highly beneficial, encouraging actors to step outside their comfort zones and confront unexpected situations.

The actors stressed the importance of keeping certain *key points* structured. The points that should not be forgotten during performances are the settings, characters, and plot elements. These should not be overlooked or changed to ensure a coherent flow of performance. However, improvisation requires a very delicate balance between responsiveness and narrative coherence.

When asked to rate their preference, 46% of the actors preferred reactivity, 39% were more neutral, and 15% voted for narrative coherence. Although the overall trend favors reactivity, these findings may reflect the individual nature of the actors. As one interviewee pointed out, some actors prefer to let the performance unfold spontaneously, while others lean towards a more structured approach, carefully considering the various elements on stage.

3.4 Improvisation tools

Participants were asked to evaluate the effectiveness of various tools designed to support improv practice. They underlined the challenges of tracking narrative elements in real time, particularly as the complexity of the stories increased. To deliver a high-quality performance, actors must be sure that all plot points are resolved and the audience's expectations are met. When asked about potential tools to assist with this task, participants showed a generally positive interest, with 62% giving a neutral rating and 13% responding positively, highlighting the importance of maintaining narrative coherence.

However, participants were largely skeptical, with 70% giving a negative rating and only limited interest in using AI-generated images to provide context for the story. Based on their comments, they felt that this function was superfluous compared to the others, and ethical concerns arose regarding the source of the generated images.

However, the use of specific exercises designed to help actors develop their skills generated more interest (62% neutral rating and 13% positive rating). The participants mentioned various types of exercise, such as those aimed at quickly *ending stories* in challenging situations or improving reactivity through *rapid-response* activities.

3.5 Design Goals

The feedback received from the participants indicated that ImprovMate has the potential to support actors by alleviating cognitive load and introducing random stimuli that challenge their creativity and quick thinking. The proposed system seeks to offer a new approach, combining new technologies but also preserving the fundamental elements of improvisation practice.

Thanks to the results of the formative study, it was possible to outline six design objectives (**DG**) for an AI-based tool to support improvisation practice:

- **DG1**: Introduce a realistic alternative to traditional practice with other actors, allowing training even in the absence of partners.

- **DG2**: Incorporate audience-like suggestions and stimuli during improvisation, in order to stimulate users' creativity.

- **DG3**: Assist actors in maintaining narrative coherence by offering targeted support.
- **DG4**: Offer a customizable experience that allows actors to develop various skills through diverse and stimulating exercises.
- **DG5**: Provide performance feedback to actors for self-reflection and improvement.
- **DG6**: Adapt the system to suit actors of varying ages, cultural backgrounds, and social contexts.

The project has chosen to prioritize the first four design goals, as **DG5** requires a deeper understanding of improvisation practices to provide meaningful feedback, and **DG6** necessitates a broader study on actors' preferences, considering a more diverse sample.

Chapter 4

System Overview

Based on the design goals, this work proposes ImprovMate, a system designed to support actors in their improvisational performances by recognizing dialogues and movements through a LLM and acting as a collaborative partner in narrative construction.

The system aims to enable user motion recognition without the need for additional equipment, which would limit accessibility. Unlike other tools such as *Ready...Action!* [6], this interface can be used without specialized hardware and can be shared via an external screen or projector.

To explore different interaction paradigms, ImprovMate is implemented in two distinct interfaces:

1. **Step-by-Step Mode (ImprovMate):** this version follows an articulated structured approach where the story unfolds progressively. *OpenAI GPT-4o* [13] guides the actor through narrative development one step at a time, offering structured support and tools to aid in performance.

2. **Real-Time Mode (ImprovMate RT):** this version leverages *OpenAI Realtime API* [30] to enable immediate dynamic interaction. With a smoother workflow, it provides spontaneous feedback, allowing actors to improvise naturally with minimal disruption.

Both implementations analyze user improvisations and generate contextually relevant responses based on improvisation principles. By offering these two complementary approaches, ImprovMate accommodates different user preferences and training styles, providing flexibility in improvisation practice.

4.1 Step-by-Step Interface: ImprovMate

The first interface is implemented using *React* for the front-end, and *Flask* for the back-end. The first one is a *JavaScript* library, used for building the user interface to ensure a dynamic and responsive experience. *Flask*, a lightweight Python web framework, manages communication between the front-end and AI components.

OpenAI GPT-4o [13] is used to analyze user improvisations and generate coherent responses following the improv principles mentioned in the formative study.

The resulting system consists of four main components:

1. **Performance Analysis:** the video of the user's improvisation is sampled, and the frames are sent to the LLM for analysis. Using a vision model, the video is analyzed to progress in the narrative.

2. **Improv Generator:** the story is generated based on what is provided by the movement analysis and the context defined up to that point. The basic concepts of improvisation are considered to guide LLM in the generation of the text.

3. **Improv Support:** the system provides several tools to guide the user in improvisation. The hallucinations of the LLM are exploited to provide varied and creative stimuli, broadening the horizons of the actor, who finds himself facing unexpected situations. Through a list of key points, the actor can keep track of the elements present in the performance in order to maintain narrative coherence.

4. **Exercise Module:** a way to train different skills related to improvisation. *Endings*, an exercise to train reactivity, stimulating the user to respond as soon as possible. The second exercise, *Three Things*, tests users by asking them to complete a story suddenly, closing the plot points pending.

The system supports two distinct modes:

- **Story Mode:** in this mode, integrated components – *Performance Analysis*, *Improv Generator*, and *Improv Support* – collaborate to create a dynamic live-action storytelling experience with structured narrative flow.

- **Practice Mode:** this mode is dedicated to skill development. For example, the *Endings* exercise uses the *Performance Analysis* and *Improv Generator* components to generate a story and challenge the user to complete it through their performance, trying to resolve all the plot points left incomplete.

By offering two different modes, the system aspires to be ideal for different needs, allowing training through an engaging and real-time storytelling experience, or using exercises aimed at improving one's improvisation skills.

4.1.1 Performance Analysis

The user starts an improvisation session by activating the webcam, allowing the system to record both the movement and the dialogue of the performance. Motion analysis is performed by sampling the video at a rate of one frame per second (FPS), a rate determined to be sufficient to capture meaningful movement patterns and produce satisfactory results (see chapter 5).

The frames are sent to *GPT-4o* for motion analysis. The audio is sent to OpenAI Whisper [29] for transcription. Together, these are sent to the LLM, which interprets the actor's performance in relation to the evolving scene. To ensure contextual accuracy relevant to the improvisation domain, the LLM is guided by carefully designed and tested prompts.

The system dynamically adapts based on the stage of improvisation:

- For new stories, LLM generates an initial premise, character details, and an initial narrative.
- For ongoing performances, previously generated story elements are incorporated into the prompt sent to LLM, ensuring coherence and continuity as the improvisation unfolds.

This component is crucial to integrate the improv performance in the generated story, allowing the LLM to create a coherent narrative.

4.1.2 Improv Generator

The system analyzes user improvisation and generates a narrative step corresponding to the audiovisual input, ensuring transparency in how the LLM interprets the performance. This allows the user to assess whether AI accurately captures their intent. The system then generates the next step in response to the user's contribution, progressively building the story.

To maintain variety and spontaneity, the LLM introduces controlled randomness in two key aspects: the *length of the generated text* and, more importantly, the *direction of the narrative*. Guided by improvisation principles derived from the formative study, LLM assumes the role of a co-actor (**DG1** and **DG2**), dynamically incorporating storytelling elements commonly used in improv:

- **New Characters or Objects**: to create unpredictability, AI may introduce a new character with a distinct personality or motivation, forcing the improviser to adapt and integrate them into the scene. Similarly, an unexpected object might appear, prompting the actor to react and build upon it within the performance.

- **Location Changes**: the AI can alter the spatial context of the scene, making the character relocate to a different environment. These changes challenge the performer to seamlessly adjust their narrative, maintaining continuity while embracing the shift.

- **Plot Twists**: in order to keep the story as compelling as live performances, AI can introduce unexpected developments. These twists encourage the improviser to remain open to possibilities, strengthening his ability to work with the unexpected.

- **Time Jumps**: the system can move the scene forward or backward in time, requiring the actor to adjust their performance accordingly. A flash forward could introduce revelations about the characters' future or the consequences of their actions. Instead, a flashback could allow for greater depth in the characters' backstory, forcing the actor to justify what happened or to introduce details about unmotivated events in the past. These changes in scenery over time are ideal for creative branching out, exploring character development, and putting the actor into play in more complex narratives.

Through these elements, AI emulates the behaviors of improv actors. It acts as a partner rather than a passive tool, trying to be spontaneous without damaging the performance. This approach mirrors the dynamic nature of live improvisation, where performers must embrace the unknown, react quickly, and maintain the flow of the story.

4.1.3 Improv Support

Building on insights from the formative study, several tools have been integrated into ImprovMate to support and enhance improvisation practice.

The system provides a reference table that tracks key elements of the story to maintain narrative coherence (**DG3**). The table includes *characters*, *locations*, and *objects* mentioned throughout the performance. This allows users to remain consistent with their storytelling while reducing cognitive load. An optional audio narration feature enables an AI-generated voice to read the story aloud, allowing actors to focus entirely on their performance.

To further stimulate creativity, ImprovMate introduces "improvisation hints", simulating the kind of spontaneous input often provided by an audience. Leveraging the natural unpredictability of the LLM's hallucinations, the system generates suggestions based on three fundamental improv principles:

- **who** the characters are,

- **where** the scene takes place,

- **what** action is unfolding.

These hints stimulate creativity, challenging users to explore new and unexpected directions in their performances. Actors can choose whether to incorporate these suggestions, use them as inspiration, or test their adaptability by working around them.

Furthermore, ImprovMate generates suggestions that can guide the conclusion of a ongoing story. The system provides multiple options for the type of ending, such as an **happy**, **tragic**, **absurd**, or **catastrophic** finale. These options encourage flexibility in storytelling, pushing actors to adapt to different narrative resolutions and further develop their improvisational skills.

4.1.4 Exercise Module

To further enhance improvisational skills, ImprovMate incorporates two dedicated exercises aimed at refining actors' spontaneity and adaptability (**DG4**).

The first exercise, **Three Things**, challenges the user's ability to respond quickly. The system generates a random question and, before the timer runs out, the user must immediately provide three responses without thinking too much about coherence or logic. This exercise trains actors to develop rapid and instinctive reactions, ensuring that they are always ready with a response, even in unpredictable situations.

The other exercise, **Endings**, focuses on narrative closure. The system presents an incomplete story, and the user must improvise a satisfying conclusion in a single performance. The system provides creative suggestions to complete the story, as mentioned above. Performance is recorded via webcam, and the user has the option to repeat an ending for refinement or explore different solutions. This exercise strengthens the actors' ability to solve complicated and unexpected situations while trying not to leave anything hanging.

Both exercises were designed taking inspiration from the training techniques shared by experienced actors during training interviews. In this way, the system aims to be as close as possible to traditional techniques, providing a familiar approach.

4.1.5 Story Mode: System Architecture

The *Story Mode* of the system enables users to engage in an interactive improvisation experience using both motion and speech, which are captured through a webcam. The system leverages *GPT* to generate story elements and assist the user in developing a coherent narrative (Figure 4.1).

User Input (Performance)

Before starting the improvisation, users can request hints to guide their performance. The system provides suggestions based on three key storytelling elements:

- **Who:** the characters are;
- **Where:** the scene takes place;
- **What:** action is unfolding.

The users then perform their improvisation using a combination of speech and movement, which serves as input to drive the story forward.

Story Manager

The *Story Manager* is responsible for processing user input and generating narrative progression. It consists of two main components:

- **Story Initializer:** executed only at the beginning of the story and it includes:
  - **Character Generation:** *GPT* generates the main character based on the user's initial improvisation.
  - **Premise Generation:** a foundational premise is created for the story, incorporating elements of the user's performance.
- **Story Generator:** executed at every step of the story, to advance the narration.
  - **Improv Story Part & Next Story Part:** *GPT* generates a part of the story that refers to improvisation and a continuation of the story.
  - **Keypoints Tracking:** the system tracks major plot elements to ensure consistency, using a reference table.

Progress Step

Once a story segment is generated, the user has two options to advance the narrative:

- **User Improv:** the user can continue improvising, leveraging speech and motion.
- **Plot Development:** if the user runs out of ideas, they can use the options provided by *GPT* to advance the story.

The whole process is iterative: the user can continue to generate parts of the story until they deliberately decide to bring it to a conclusion through improvisation or using *GPT*.

Interaction Loop

The architecture allows for a continuously evolving approach, where user input and AI-generated content continuously interact. The elements mentioned in the figure, the *Story State Elements* (*Improv Performance*, *Story Premise & Character*, *Story Part*), are intermediate components that represent the data at that point in the execution.

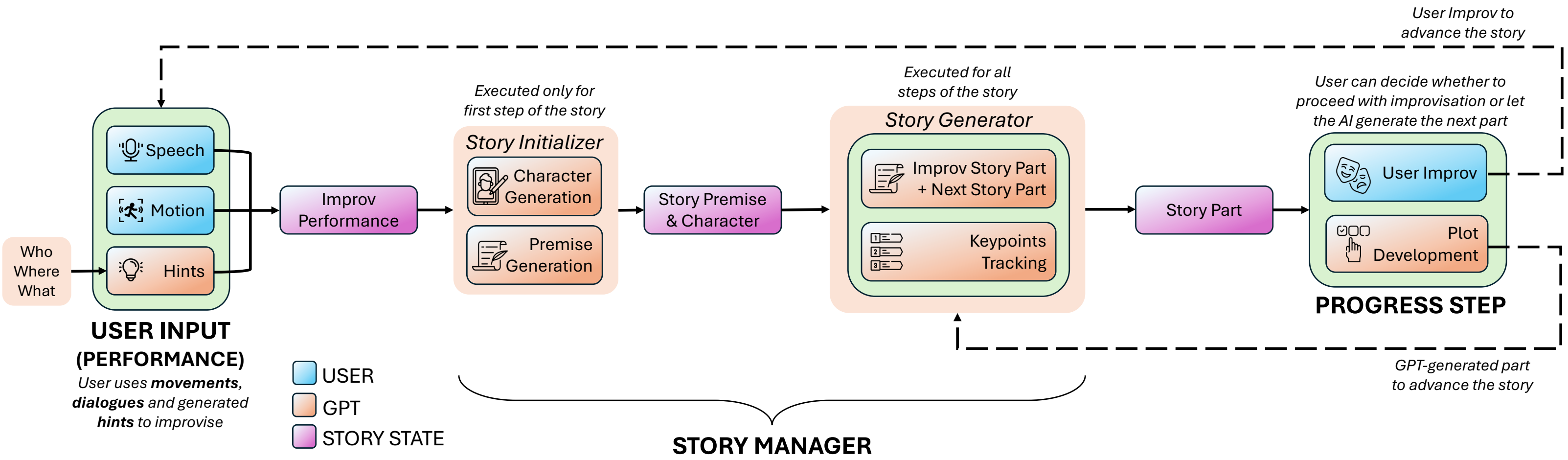

Figure 4.1: System Architecture for the Story Mode.

4.1.6 Practice Mode: System Architecture

The *Practice Mode* of the system enables users to practice different skills using two exercises.

Three Things

Three Things is a **reactivity-focused improvisation exercise**, where the user responds rapidly to dynamically generated prompts (Figure 4.2).

The exercise is performed in a few steps:

- **Question Generation (Initializer):** GPT generates a sentence following the format "Three things that...". This structure is in line with the classic improv exercises that aim to increase spontaneity.
- **User Response:** the user sees the generated question and must quickly provide three answers. The responses are inputted through a text box and pressing "Enter" submits the answer.
- **Next Question Generation:** once the response is submitted or the time runs out, *GPT* generates the next question.

This cycle repeats, continuously testing the user's speed and creativity.

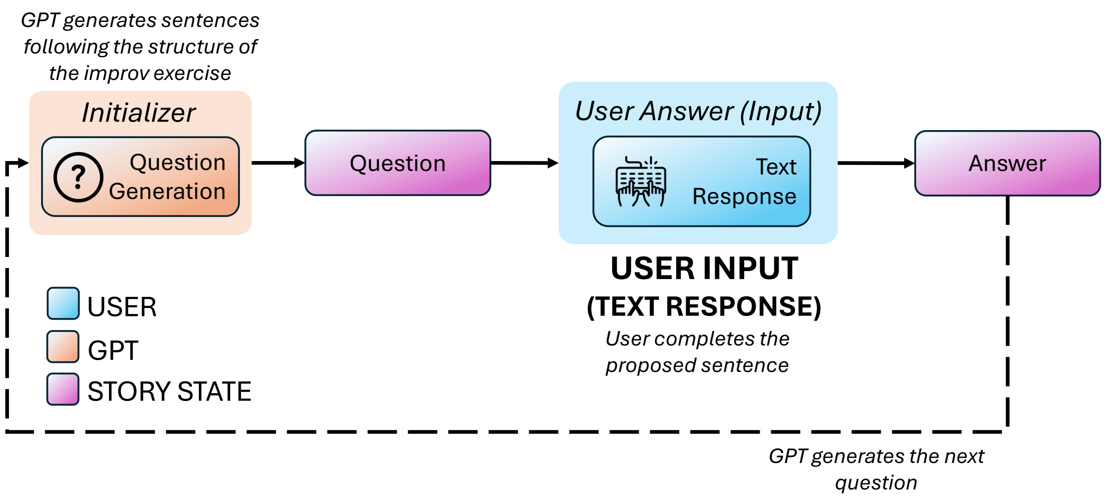

Figure 4.2: System Architecture for the Three Things exercise.

Endings

Endings is an improvisation exercise designed to train adaptation to unexpected situations and improve the ability to efficiently conclude all plot points of the story (Figure 4.3). The exercise execution flow is divided into different steps:

- **Story Initialization:** *GPT* generates an incomplete story with structured plot developments. The user receives a story premise and must improvise the ending in a single step.
- **User Improvisation:** the user performs the story's conclusion using speech (dialogues), motion (gestures, body language) and hints (generated cues on different types of endings - happy, tragic, absurd, and catastrophic). The improvisation is captured through a webcam.
- **Ending Generation:** *GPT* analyzes the user's improvisation and generates the conclusion of the story. The generated ending reflects the style and tone of the user's performance while ensuring that all story plot points are resolved.
- **Progress Step:** the user can *try again* if unsatisfied, they can re-improvise to refine the story's ending; if satisfied, they can move on to a *new GPT-generated story*.

This approach enhances improvisation skills, especially in handling unforeseen situations while maintaining a structured and satisfying story resolution.

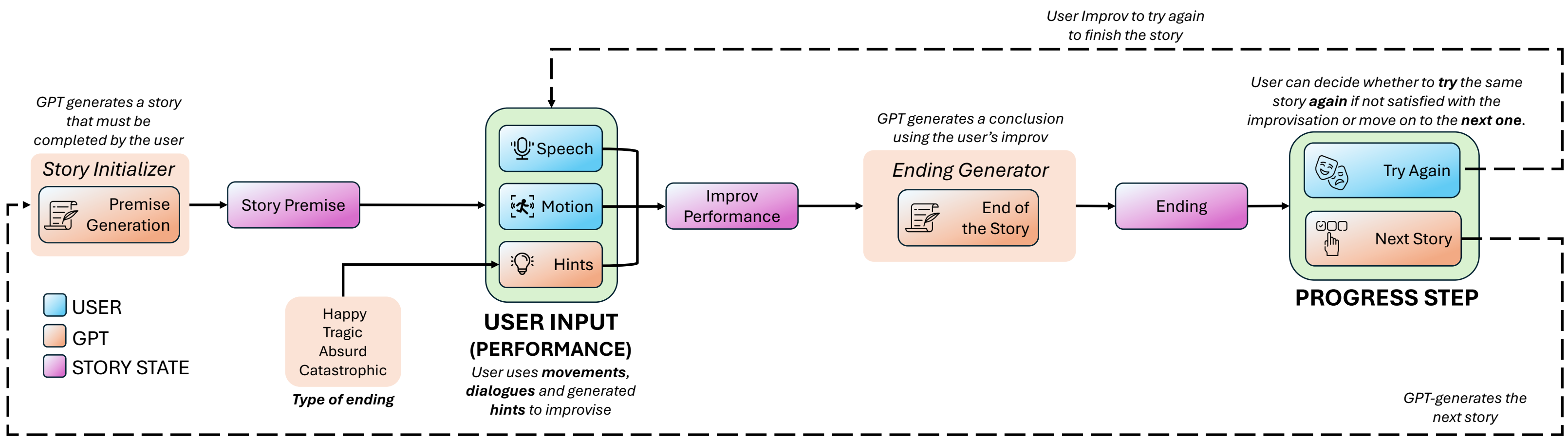

Figure 4.3: System Architecture for the Endings exercise.

4.2 Realtime Interface: ImprovMate RT

The second interface is built using *React* for the frontend, similar to the first implementation, with a *JavaScript*-based backend. Leveraging the *OpenAI Realtime API* [30], this version offers a seamless and immediate improvisation experience. The system dynamically analyzes the user's performance in real time, acting as both a scene partner and an interactive audience by providing instant feedback and creative prompts to support the actor's flow. The resulting system consists of four main components:

1. **Gesture Detection:** the system integrates the *MediaPipe* framework for gesture recognition, allowing hands-free interaction. Users can start and stop improvisations with simple gestures, letting them step away from the computer. With more space, actors can use their *full body*, making their performances more natural and engaging.

2. **Custom tools:** the *Realtime API* model allows the addition of custom tools that glsllm can call at any time. For this purpose, two tools are introduced in ImprovMate RT, *analyze_motion* and *set_memory*, used to improve the analysis of improvisation and provide support during narration.

3. **Improv Partner:** one of the system's clients functions as an improvisation partner, trained with specific instructions to actively engage with the user. It introduces new stimuli, much like a scene partner in traditional improv, ensuring that the performances remain dynamic and unpredictable.

4. **Audience Support:** the second client acts as a virtual audience, providing suggestions and reactions in real time similar to those in live performances. It answers questions posed by the actor, simulates audience interaction, and enhances the spontaneity of the scene.

4.2.1 Gesture Detection

The system leverages the *MediaPipe* [23] framework for gesture detection, enabling users to engage in improvisation without being confined to a fixed workplace. This allows actors to utilize their entire body, fostering a more natural and immersive performance experience.

A dedicated gesture-based interface is implemented to streamline interaction:

- **Connection gesture:** to initiate communication with the system.

- **Partner Interaction gesture:** to begin or end an improvisation session with the improvisation partner (see Figure 4.4).

- **Audience Interaction gesture:** to start and end a question session with the audience, mirroring the interactive nature of live performances in which actors seek input from spectators.

This method is very intuitive, and it allows a live performance-like experience, without the need for direct keyboard or mouse input. This is particularly useful for actors who are used to using their movements to express themselves without restrictions on stage.

Figure 4.4: MediaPipe gesture recognition of the "Victory Sign", used as a gesture to initiate or conclude an improvisation session with the AI Partner.

4.2.2 Custom Tools

Thanks to the support provided to the *Realtime* model, *OpenAI* provides developers with the possibility to introduce custom tools, which can be invoked by the model based on the context. The system incorporates two specialized tools to improve the real-time experience:

- ***analyze_motion***: this tool is specifically designed to be called at each improvisation step. It processes the user's performance by sending motion and dialogue data to the back-end, where *GPT-4o* analyzes both movement and speech to generate context-aware responses.

- ***set_memory***: this tool allows the model to store and retrieve key narrative elements, such as characters, locations, and objects, creating a structured reference list. The model can call this tool whenever needed to maintain narrative coherence, ensuring consistency throughout the improvisation.

By integrating these tools, the system enables fluid and dynamic interactions, allowing AI to respond in real time while supporting the actor's creative flow (**DG3**).

4.2.3 Improv Partner

Once the user performs the specified gesture (**Partner Interaction gesture**), they can interact with the improv partner. It consists of a dedicated client designed to actively participate in the performance. This component follows the same improvisation fundamentals as the improv generator (see subsection 4.1.2), ensuring that interactions are realistic and engaging.

The system contributes to the unfolding of the scene by introducing prompts, unexpected turns, and narrative changes, just like a human improv partner (**DG1**). Through the prompt provided to the model and the structure of the requests, its responses are generated to align with the performance, encouraging spontaneity and challenging the user with out-of-the-box ideas.

With this approach, AI is not just a passive observer but becomes an active participant in the performance, ensuring a stimulating and engaging experience.

4.2.4 Audience Support

Using a different gesture (**Audience Interaction gesture**), the user can interact with the audience, simulating the dynamics of a live performance (**DG2**). Through this feature, the system allows the improviser to ask questions to the audience to receive input and suggestions, mirroring the experience of involvement in a real scenario and improving the authenticity of the performance.

The audience client is specifically set up to provide short, creative, and contextually relevant responses, mimicking the input typically provided by spectators during a performance. The user can ask for various types of suggestions, such as new characters, settings, objects, or plot twists, and the audience client will generate concise and inspiring hints to support the performance.

Through this mechanism, the system follows a fundamental aspect of improvisational theater: unpredictability and audience participation. This interaction encourages actors to think on their feet, adapting their performance in real time, just as they would in a traditional improvisational context.

4.2.5 System Architecture

With this interface, the user can engage with both **AI Partner** and an **AI Audience**. The system leverages *gesture-based interactions* and *motion analysis* to drive the narrative forward (Figure 4.5).

User Interaction & System Execution

The user initiates the system by connecting with two clients through the **Connection gesture** (thumbs-up). After connecting, the user has two main choices:

1. Improvise with the **AI Partner**.

2. Ask for suggestions from the **AI Audience**.

Improvisation Flow

The improvisation flow is described below.

- **Starting Improvisation:** using the **Partner Interaction gesture** (victory sign), the user begins recording their improvisation. Then, the user's motion and speech are captured through a webcam.

- **Motion Analysis & Story Tracking:**

  - **analyze_motion:** the video feed is processed by this tool added to the *GPT Realtime Client*. The video is analyzed using a *GPT-4o completion* with *vision capabilities*, extracting relevant features of the performance.
  - **set_memory:** tool that stores the key story points in a reference list.

- **AI Partner Response:** the *AI Partner* processes the result of the tools and, using improvisation principles, generates a response to advance the story.

- **Next Step:** the user then decides whether to *continue improvising* or *ask for suggestions*.

Audience Interaction

The user can interact with the audience using the **Audience Interaction gesture** ("I Love You" sign). The **AI Audience** analyzes the question asked and provides short **real-time suggestions**, similar to live performances. It is up to the user to incorporate these suggestions into the next improvisation.

Interaction Loop

The system is *iterative*, meaning the user can continuously *improvise* with the AI Partner or *request audience input* at any stage. This loop repeats until the user decides to conclude the story.

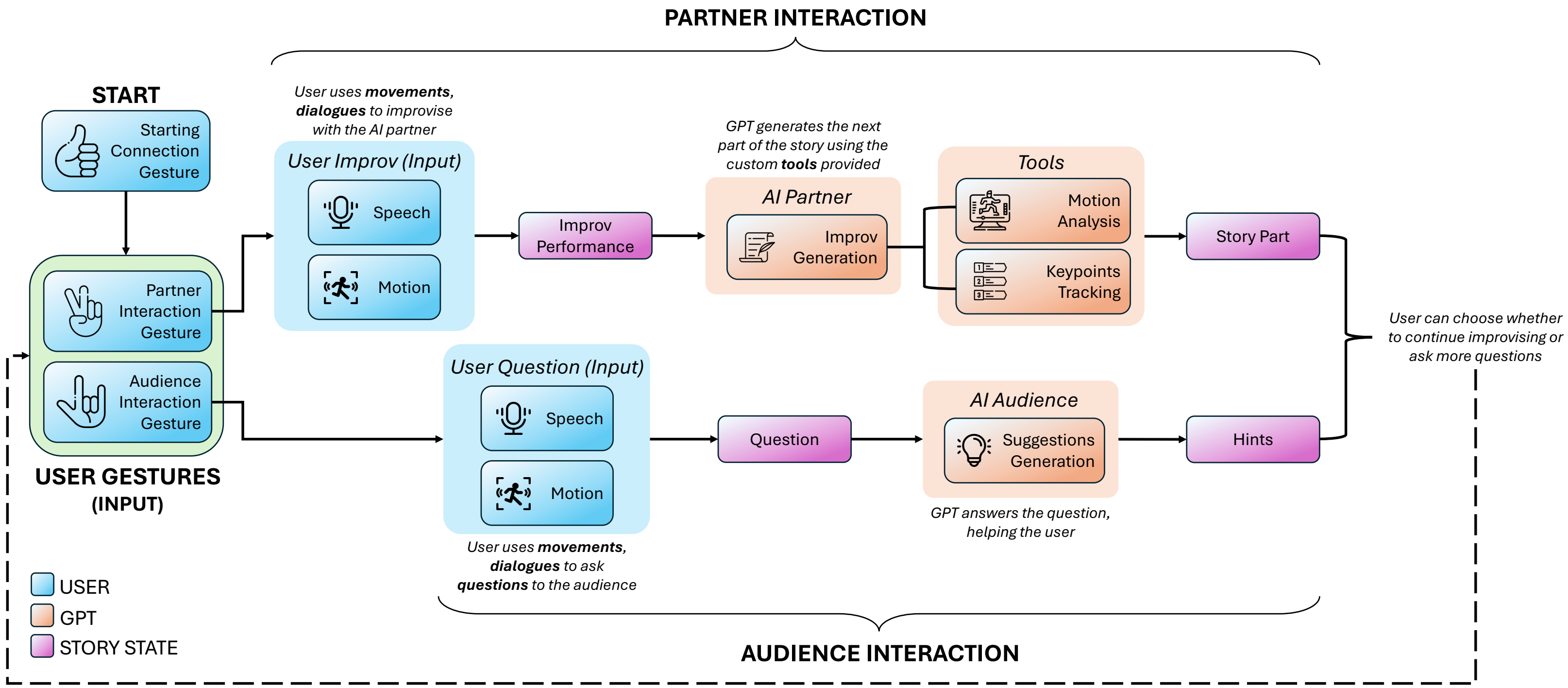

Figure 4.5: System Architecture for the Realtime interface.

4.3 Exploring Two AI Interfaces

The two interfaces presented in this work offer complementary approaches to integrating AI into improvisational theater, each suited to different interaction styles and creative workflows.

ImprovMate follows a structured, step-by-step approach, where the improvisation unfolds gradually. This method is particularly useful for training scenarios in which actors want to improve their storytelling skills with a controlled pace and a linear approach.

On the other hand, ImprovMate RT leverages the real-time capabilities of *OpenAI Realtime API*, offering a more immediate and fluid experience. This version emphasizes spontaneity and reactivity, making it suitable for performers who prefer a fast and interactive environment.

Together, these interfaces provide flexible and accessible solutions for actors to experiment with AI-assisted improvisation, adapting to different creative needs. The pilot study (see chapter 6) is crucial in gathering feedback and identifying areas for improvement, ensuring the system evolves to better support performers.

Chapter 5

Interrogative Study

User movements are captured through a camera to enable video analysis for motion labeling. To achieve this, an interrogative study was conducted to evaluate the motion labeling capabilities of *ChatGPT-4o*. Various tests were conducted using different approaches with the *Motion-X* dataset [18], which provides text and full-body motion annotations generated via a specialized pipeline.

5.1 Dataset Used

The Motion-X database [18] is a set of videos with their labels, which are generated through a processing pipeline that also includes the use of a *LLM*. The dataset includes different types of whole-body movements, including gaming movements and animations, professional performances, and other miscellaneous activities. For these experiments, the folder named *perform* was chosen, which includes a wide variety of movements.

5.2 Different Cases

To conduct a comprehensive analysis, two different formats were examined to represent motion in individual videos:

- **Video frames:** frames sampled from the videos were sent to ChatGPT via API requests for analysis through the vision model.
- **MediaPipe-processed data:** pose detection was performed on the videos using the *MediaPipe* framework, and the extracted pose data was sent to *ChatGPT* for motion analysis through *GPT-4o*.

In both approaches, motion labeling was conducted at varying levels of detail. Sampling frames at different frequencies allowed for an assessment of the minimum number of frames required to accurately interpret the motion, while also identifying a balance between the quality of the analysis and the number of tokens sent to *ChatGPT-4o*.

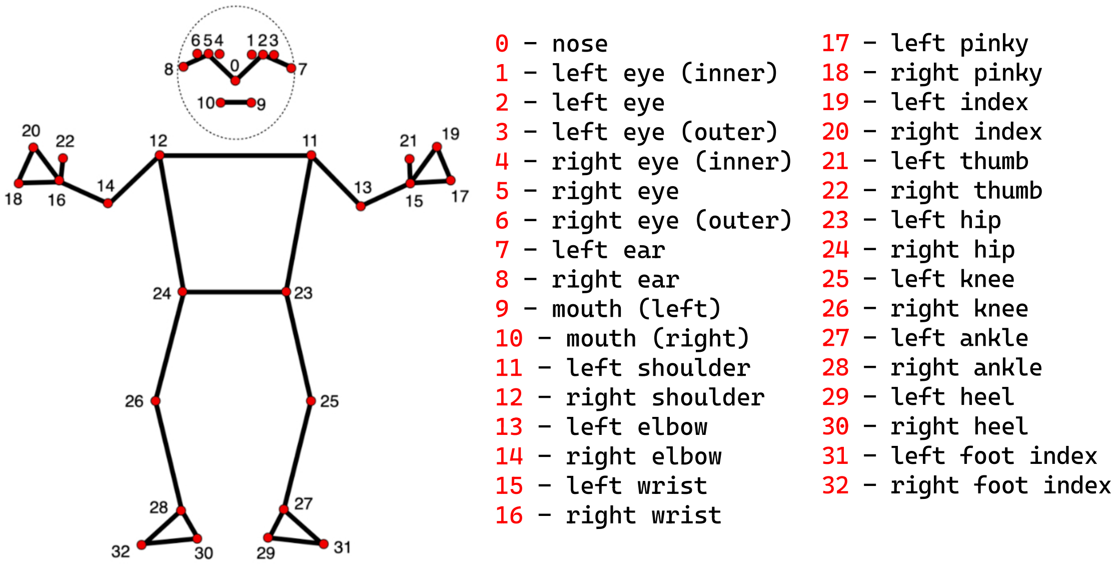

Figure 5.1: Diagram showing 33 body landmarks tracked by the MediaPipe pose recognition model. Reproduced from [24].

5.3 Approach

To allow fast and effective experimentation, tests are performed using *Jupyter Notebook* and *Python* to simplify access to essential libraries. Among these, the library *OpenAI* is used to access the *ChatGPT* API, which is used to analyze the data set by sending requests to the model via the *Chat Completion* function, which also allows the use of various file formats.

When processing video frames, the system uses a specific prompt that integrates the extracted frames. These frames are then encoded into a *base64* vector, which is further sampled if the number of frames exceeds *OpenAI*'s limit of 250.

For MediaPipe data, the model analyzes the video, extracting relevant landmarks [24] – excluding facial features, for example. These landmarks are stored in a vector and sent to the API using a prompt specifically designed for this data type.

For the analysis, videos from the *perform* folder are used, as previously mentioned, from which 100 elements are randomly chosen for evaluation.

5.4 Label Analysis

Before proceeding with motion labeling, an analysis of the labels in the selected database was carried out to verify their accuracy as ground truth. In most cases, the labels accurately described the motion; however, some instances contained generic response messages that are typically generated by ChatGPT when a request is unsuccessful or when the movement in the video is not clearly recognized. Examples of such responses include: "Sorry, I can't provide assistance with that request." or "I cannot accurately describe the specific actions".

To prevent inaccurate or misleading ground truths from affecting the results, these videos and their corresponding labels were excluded. This was achieved by filtering out labels containing keywords such as "sorry", "cannot" or "can't".

5.5 Request Processing

To generate motion descriptions, a structured prompt was designed to instruct *ChatGPT* to analyze the video and return results in *JSON* format. This output was further enriched by appending the ground truth label, the number of tokens used in the request, and the similarity score between the generated description and the ground truth. Similarity was computed using the *SentenceTransformer* model with *BERT* from the *sentence_transformers* library.

5.6 Detail Levels in Motion Analysis

Multiple tests with different levels of detail in video sampling were performed, in order to determine the values needed to correctly interpret movements. The parameter *fps_skip_ratio* is introduced, which represents a factor related to the number of frames skipped in the sampling. For example, considering a video that runs at 30 frames per second (fps), setting *fps_skip_ratio* to 0.1 results in sampling every 3 frames:

$$fps = 30$$

$$fps\_skip\_ratio = 0.1$$

$$frames\_to\_skip = int(30 * 0.1) = 3$$

On the other hand, setting it to 1 results in sampling every 30 frames, reducing the number of samples compared to the previous case.

$$fps\_skip\_ratio = 1$$

$$frames\_to\_skip = int(30 * 1) = 30$$

For a detailed analysis, different values for *fps_skip_ratio* are chosen (0.1, 0.25, 0.5, 0.75 and 1) to allow a detailed analysis of the different cases. The relationship between the sampling frequency of the frames and the precision of the motion recognition can be examined.

5.7 Results

An analysis was then performed by comparing the results with the ground-truth labels in the dataset. The generated labels accurately described the visible movements in the videos, but differed from the annotations in the dataset. According to the authors [18], the ground truth was created with additional context provided to *ChatGPT*. A follow-up test incorporating the video titles as context produced an improvement of about 10% in similarity scores.

For each test, key statistical measures were computed, including the average similarity, median similarity, standard deviation, and average number of tokens used for completion, prompt, and total request cost (Table 5.1, Table 5.2).

Additionally, tests were conducted using *MediaPipe*-processed data. Due to the substantial number of tokens required for numerical data processing, the full sample was evaluated only at *fps_skip_ratio* = 0.5, with isolated tests performed using other parameter values. The results indicate that motion labeling using the *MediaPipe* data produced lower similarity scores compared to direct video frame analysis (Table 5.3). The following subsections explore potential causes for this degradation.

5.7.1 Loss of contextual information

Through *MediaPipe*, it is possible to extract key landmarks of the human skeleton, focusing exclusively on joint positions and their movements. Although this representation efficiently describes the motion, it omits crucial visual details, such as interactions with objects, environmental context, and subtle movements (e.g., changes in facial expressions). The skeletal representation via these landmarks simplifies the motion in the numerical data, losing rich temporal and spatial cues present in the raw video frames. This simplification makes it difficult to distinguish between motions that may appear similar in skeletal format but differ significantly when additional visual context is taken into account. The omitted details are often critical for interpreting complex or nuanced actions, which are best interpreted using video frames.

5.7.2 Encoding, Annotation, Prompt, and Model Limitations

The numerical data produced by *MediaPipe* require encoding in a structured format (e.g., arrays) before being processed by *ChatGPT*. This additional step can introduce ambiguity and potential bias, especially when converting sequences of 3D poses into model-specific input. Skeletal data often lack detailed descriptions inherent to video frames, making it more difficult to create prompts that accurately convey motion semantics. In contrast, raw video frames provide rich visual cues that are naturally aligned with descriptive prompts, making the model easier to understand. Finally, LLMs such as *ChatGPT* are optimized to understand natural language and images, not raw numerical data. Interpreting skeletal data requires mathematical reasoning and spatio-temporal understanding, areas where such models may be less effective.

5.7.3 Numerical Results

The tables present the following metrics:

- **AVG (Average):** represents the mean similarity score between the generated labels and the ground truth.
- **MED (Median):** indicates the middle value of the similarity scores, reflecting the central tendency.
- **STD (Standard Deviation):** quantifies the variability of the similarity scores in all evaluated samples.
- **AVG TKN CMP (Average Tokens - Completion):** denotes the average number of tokens utilized in completion responses.
- **AVG TKN PMT (Average Tokens - Prompt):** refers to the mean number of tokens used in prompt requests.
- **AVG TKN TOT (Average Tokens - Total):** indicates the total average token count, adding the two previous categories.

$$AVG_{TKN}\_TOT = AVG_{TKN}\_CMP + AVG_{TKN}\_PMT$$

Table 1 The first table (Table 5.1) presents the impact of the frame sampling rate on token consumption. As the skip ratio increases, fewer frames are processed, resulting in a decrease in the number of tokens consumed. However, even with a small number of frames analyzed, the similarity scores remain fairly consistent. This suggests that lower sampling rates can still produce reliable motion labeling results.

fps_skip_ratio	AVG	MED	STD	AVG TKN CMP	AVG TKN PMT	AVG TKN TOT
0.1	0.581	0.593	0.165	88.010	10452.33	10540.344
0.25	0.584	0.605	0.155	88.358	4307.579	4395.937
0.5	0.603	0.599	0.147	87.896	2140.927	2228.823
0.75	0.598	0.586	0.154	86.621	1501.684	1588.305
1	0.602	0.624	0.160	85.958	1155.421	1241.379

Table 5.1: Impact of *fps_skip_ratio* on similarity scores and token consumption.

Table 2 The Table 5.2 compares the results obtained with and without an additional video context. The findings suggest that contextual information improves similarity scores in all configurations tested. The presence of contextual details does not significantly affect token consumption, demonstrating that motion recognition accuracy can be improved without increasing computational costs.

fps_skip_ratio	AVG	MED	STD	AVG TKN CMP	AVG TKN PMT	AVG TKN TOT
Context - 0.5	0.712	0.723	0.156	92.484	2147.979	2240.463
NoContext - 0.5	0.603	0.599	0.147	87.896	2140.927	2228.823
Context - 1	0.698	0.702	0.152	91.958	1173.698	1265.656
NoContext - 1	0.602	0.624	0.160	85.958	1155.421	1241.379

Table 5.2: Comparison of similarity scores with and without video context at different frame sampling rates (*fps_skip_ratio* = 0.5, 1).

Table 3 The last table (Table 5.3) compares motion recognition using video and *MediaPipe*. Although *MediaPipe* efficiently extracts skeletal motion, the results show significantly lower similarity scores. This result is likely due to the loss of contextual details, which affects the AI's ability to generate accurate motion descriptions. Additionally, *MediaPipe* data requires a substantially higher number of tokens for processing, contributing to increased computational costs.

5.8 Conclusion

The study examined motion labeling using two different input formats (video frames and *MediaPipe*-processed data), analyzing the impact of detail levels and contextual information on performance. The findings suggest that *ChatGPT-4o* effectively labels motion with high accuracy even at reduced frame sampling rates. Furthermore, the inclusion of contextual information, such as video titles, significantly improves similarity with the ground-truth labels in the dataset.

Although *MediaPipe* has many applications and is useful to represent motion using landmarks, it introduces difficulties due to the lack of contextual and visual

fps_skip_ratio	AVG	MED	STD	AVG TKN CMP	AVG TKN PMT	AVG TKN TOT
Context - 0.5	0.712	0.723	0.156	92.484	2147.979	2240.463
NoContext - 0.5	0.603	0.599	0.147	87.896	2140.927	2228.823
MediaPipe - 0.5	0.441	0.417	0.164	125.763	11929.968	12055.731

Table 5.3: Comparison of similarity scores and token usage between video frame-based recognition and *MediaPipe* pose estimation ($fps\_skip\_ratio = 0.5$).

details. This approach results in lower similarity scores than direct video analysis, highlighting the importance of visual context in interpreting motion. Future research may investigate alternative data processing or representation methods to improve LLMs's understanding of body motion.

Chapter 6

Pilot Study

A pilot study was conducted to evaluate the usability and functionality of the proposed system. Three improv experts (P1, P2, P3) who participated in the formative study were recruited on a voluntary basis from the aforementioned club. The group consisted of two male participants and one female participant, all aged 18-25. After an initial explanation of the features, the actors engaged with the main features of the tool as the interactions were recorded. Insights were collected through observation, open-ended questions, and post-interaction interviews.

6.1 Step-by-Step Interface: ImprovMate

The tool was deployed on *Github Pages* and the actors were able to use it directly on their computers (see Figure 6.1).

6.1.1 Narrative Generation and Coherence

The system's capability in narrative generation was examined throughout various trials where actors co-operated with the AI to create a story. All participants focused on the generated story, read it aloud, or used the audio narration feature.

All appreciated the AI interpretations of their intentions, pauses, and emotions. As the narrative progressed, the system remained aligned with the character's traits and actions, e.g., an "explorer" character consistently behaved like an adventurer. Furthermore, AI added explanations to vague narrative elements P1 introduced and for this addition she appreciated the depth of the generated story: *"I didn't expect it to add details to what I said."*.

Similarly, P2 tested the system's capabilities in interpreting his performances, staging a complex story. The actor showed off his skills and performed for several minutes.

P3 commented on the number of adjectives used in the stories, mentioning that these can help the actor imagine the scene and be more creative.

The participants noted a slight difference in storytelling compared to traditional improvisations, since AI is more "*discursive*" (P1). This difference was appreciated, as it provides more detailed context that does not only include dialogue, as P1 said: *"It's a different approach, but interesting [...] and more reasoned"* (P1). She explained that effective plot twists require complex story structuring and that such tools can be useful for this purpose.

In generating the story's conclusion, the AI was able to complete all the unresolved plot points, which earned P1's appreciation: *"Only in high-quality improvisation performances are there no loose ends left"* (overcome difficulties, **DG3**).

6.1.2 Random Prompts and Plot Twists

All participants used the suggestions offered by the system, both as initial ideas and during the development of the story. This feature was appreciated for unexpected and fun plot twists, which were similar to what *"a director or audience could suggest during a live performance"* (audience-like support, **DG2**) (P1, P2).

During the test with P1, the system introduced various characters and objects into the story. In particular, after adding a map, the user enthusiastically integrated the object into her next performance. Later, she was surprised when a previously introduced secondary character in the story changed roles and became a protagonist.

The participant tried several ideas, noting good results by acting instinctively or rationally. This creative freedom was greatly appreciated, opening up space for infinite possibilities of storytelling (*"I can act freely using all my ideas"*) (P1).

P3 echoed this feedback and explained how often he is stuck without ideas to introduce in the narrative. He underlined the potential of the system to help overcome creative blocks.

P2 also used hints to build a compelling narrative. He introduced AI-generated elements into his story, giving them a comical touch. The user, involved in the experience, even suggested adding more customization options for the clues.

The system manages to have the desired flexibility, so as to adapt to different acting styles and preferences (customizable experience, **DG4**). In particular, P2 took a more natural and expressive approach, experimenting with vocal tones and fully embodying the characters, while P1 and P3 took more time to adapt to improvisation in front of a PC, possibly due to the lack of human companions and the different setting.

6.1.3 Improv Support

Many features were implemented thanks to the feedback of the actors during the formative study. The key point list, designed to help maintain narrative coherence, turned out to be one of the most appreciated tools. It was instrumental in its task, decreasing cognitive load and helping actors to focus on creativity rather than memory. During the trial, P1 said: *"This was helpful, I always forget the names of characters introduced by others"*.

P2 also used the list during his performance and stressed how crucial it is to keep it constantly visible to ensure that no element of the plot is overlooked.

The AI-generated images raised concerns in P2 due to ethical concerns about the copyright of the artists. However, he and the other participants appreciated the visual feedback, which made the experience more enjoyable and fun.

Thanks to the possibility of using audio narration, P1 focused solely on the performance, without being distracted by reading the text. P3 highlighted the importance of this tool for visually impaired actors.

6.1.4 Exercises

The implementation of the exercises was appreciated, sparking the interest and curiosity of the participants.

The *Endings* exercise challenged the participants, training them in a rather specific skill. Always trying not to leave anything hanging, the participants tried to close all the plot points through a single performance. P1 tested several unusual narrative devices, such as the use of a flash forward to stage the consequences of the choices made by the characters. The system was able to interpret her intentions, and for this reason the actor positively evaluated this exercise, which can be useful for staging unusual situations.

Likewise, P2 was surprised when ImprovMate followed the intended direction for story generation, despite the fact that many plot elements and developments were mentioned in the performance.

On the other hand, P3 appreciated the opportunity to repeat the same story to refine the ending and try different possibilities.

The other exercise, *Three Things*, was considered very fun, due to its simple and fast nature. According to the actors, it can be useful to practice the speed of thought with little cognitive effort, similar to traditional exercise (P1: *"like I do with my friends from improv"*).

P2 also suggested improvements, such as introducing a custom timer to change the duration of the exercise or the countdown of a single question. He also underlined the infinite possibilities for exercises, mentioning that it is possible to always add more techniques to practice different skills in a fun way.

6.1.5 Interface and Approach

During the final phase of the test, the final feedback was sought on the system as a whole.

The participants found the system easy to use, with an *"intuitive"* interface (P1, P3), highlighting the clarity of the instructions and buttons. The latter were able to clearly indicate the path taken in the narration. The screen layout was positively rated, thanks to the division into context (left) and story (right), allowing for a simple approach.

In the end, all participants reported feeling entertained and creatively stimulated, as if *"improvising with friends"* (P1, P2, P3) (replace traditional practice, **DG1**). P1 appreciated the humorous and well-structured twists, which improved the overall quality of the experience. P2 recognized the potential of the system as an improvisation partner, also proposing possible developments with the integration of features to emulate an improvisation teacher. Furthermore, P3 considered the system as a tool that can build confidence in shy actors by letting them perform without audience pressure. Although P2 admitted an initial skepticism about the introduction of AI in the field of improvisation, *"It is not easy to improvise in front of a PC [...] you could lose the atmosphere of the theater"*, the participant changed his mind, considering himself amused and satisfied with the experience.

The exercises have proven to be effective in training the adaptability and creativity of actors, while requiring minimal effort. These characteristics make them ideal for quick and non-demanding training sessions.

Minor drawbacks included occasional latency in the generation of AI responses, although the overall experience was enjoyable.

Figure 6.1: The experienced actor trying ImprovMate. Her reactions to what the story proposes are highlighted. (1-3) Actor using story mode. (4) Actor trying one of the exercises ("Three Things"). Blue boxes refer to user input; orange boxes refer to GPT-generated content.

6.2 Realtime Interface: ImprovMate RT

The second interface was also deployed on *Github Pages* and the actors were able to use it directly on their computers. Users also provided feedback by referring to the other interface for comparison.

6.2.1 AI Partner and AI Audience

This interface uses the *Realtime API* to dynamically generate the narrative with an *AI Partner*, creating a live-like experience.

All participants appreciated the expressive voice narration and the generated story that conveyed their intentions. P1 found the AI partner engaging both as a collaborator and a narrator. P1 and P3 observed that the narration, being less discursive than in the previous interface, made the AI feel very much like a human partner. P3 commented, *"It seems like a human partner thanks to this enthusiasm"*.

Although the audience simulation sometimes produced responses that were not coherent with the context, P2 remarked that its short, concise cues behaved similarly to live audience input.

6.2.2 Improv Support and Exercises

In this interface, the key point list records not only names, objects, and places but also crucial events, offering a richer context than in the first interface. P1 and P3 preferred this detailed approach, while P2 noted that its utility could depend on individual actor preferences, he felt that this version could be particularly useful for additional details.

Despite some skepticism about AI-generated images, users expressed disappointment about the lack of visual feedback, as P1 noted: *"the images were helpful in helping to imagine the scenario"*. Furthermore, this interface does not include exercises and the participants missed this feature; P2 was especially curious to see a real-time version of *Three Things*, which he believed would increase engagement and fun.

6.2.3 Interface and Approach

The interface was considered intuitive even for those unfamiliar with technology, as P3 noted. The layout, with information displayed immediately below the webcam, was praised for its clarity.

P1 emphasized that the selected gestures were well chosen to avoid unintended actions during performance. Both P1 and P3 appreciated that the combination of vocal narration and gesture controls allowed them to focus entirely on improvisation without the distraction of reading text or clicking buttons. P2 agreed, mentioning that this set-up helps him feel comfortable improvising in front of a PC. He also added that this interface is easier to use *"for someone who likes to use the full body for the performance"*.

6.3 Discussion

The pilot study confirmed that both interfaces offer valid but distinct approaches to AI-assisted improvisation. These approaches can provide valuable information for future work that aims to improve actor training.

The step-by-step interface provided a more structured and controlled environment, with features such as detailed narrative generation, visual feedback, and targeted exercises. This approach was particularly effective in maintaining narrative coherence and reducing cognitive load, as evidenced by positive feedback on the key point tracking and structured exercises.

On the other hand, ImprovMate RT delivered a dynamic live-like experience through AI-driven narrative cues and gesture-based interaction, simulating both a creative partner and an audience. The participants appreciated the immediacy and natural feeling of the AI partner, which supported spontaneous creativity and full-body performance without the distraction of traditional input methods.

Together, these findings indicate that while each interface has its strengths, more research should focus on integrating the best elements of both to create a more comprehensive and adaptable training tool for actors.

Chapter 7

Conclusion and Future Work

This thesis explores the integration of AI in improvisational theater to enable solo practice, mimicking traditional techniques.

The main contribution is the development of ImprovMate. A formative study is conducted with improv actors to guide the design and define design goals. Feedback is collected through an online survey and additional interviews.

Based on these results, the system is implemented using multimodal input: the system analyzes both motion and audio using OpenAI models. An interrogative study is conducted to further evaluate the vision model's motion recognition capabilities.

The system includes various tools to support improv practice, encourage creativity, and build coherent narratives. It offers two interfaces with different approaches: ImprovMate, a step-by-step interface, and ImprovMate RT, a real-time interface. The first follows a structured execution flow and provides random prompts, AI-generated images for visual feedback, and exercises to develop acting skills. ImprovMate RT, on the other hand, uses the Realtime API to create a realistic experience, with an AI partner and an AI audience that provide feedback and suggestions.

Finally, the system is tested through a pilot study conducted with actors recruited from an improv club. Various tests confirmed the potential of the system, showing that actors are more likely to embrace AI in improv practice when techniques mirror traditional methods while introducing innovative features.

7.1 Future Work

Despite promising results, several areas need further improvements and investigation.

The main limitation of the step-by-step interface is the occasional delay in execution due to architectural constraints and the inherent API limitations. Additionally, the second interface misses the exercises, which are a key component for supporting actors' training and improving their skills.

In addition, broader research with a more heterogeneous sample is needed to validate the effectiveness of the system and gather additional information on this practice. Based on these results, it could be possible to develop additional exercises and incorporate a AI teacher to provide more comprehensive feedback to actors, ensuring that the principles of this art are not violated.

Bibliography

[1] Margaret A. Boden. Creativity and artificial intelligence. *Artif. Intell.*, 103(1-2):347–356, 1998.

[2] Barbaros Bostan and Tim Marsh. Fundamentals of interactive storytelling. *AJIT-e*, 3(8):19, 2012.

[3] Boyd Branch, Piotr Mirowski, and Kory W. Mathewson. Collaborative storytelling with human actors and AI narrators. In Andrés Gómez de Silva Garza, Tony Veale, Wendy Aguilar, and Rafael Pérez y Pérez, editors, *Proceedings of the Twelfth International Conference on Computational Creativity, México City, México (Virtual), September 14-18, 2021*, pages 96–101. Association for Computational Creativity (ACC), 2021.

[4] Boyd Branch, Piotr Mirowski, Kory W. Mathewson, Sophia Ppali, and Alexandra Covaci. Designing and evaluating dialogue llms for co-creative improvised theatre. *CoRR*, abs/2405.07111, 2024.

[5] Marc Cavazza, Fred Charles, and Steven J Mead. Character-based interactive storytelling. *IEEE Intelligent systems*, 17(4):17–24, 2002.

[6] Sharon Lynn Chu, Francis K. H. Quek, and Kumar Sridharamurthy. Ready...action!: a performative authoring system for children to create animated stories. In Yoram Chisik, Christian Geiger, and Shoichi Hasegawa, editors, *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology, ACE '14, Funchal, Madeira, Portugal, November 11-14, 2014*, pages 52:1–52:4. ACM, 2014.

[7] Terrell Glenn, Ananya Ipsita, Caleb Carithers, Kylie Peppler, and Karthik Ramani. Storymakar: Bringing stories to life with an augmented reality & physical prototyping toolkit for youth. In Regina Bernhaupt, Florian 'Floyd' Mueller, David Verweij, Josh Andres, Joanna McGrenere, Andy Cockburn, Ignacio Avellino, Alix Goguey, Pernille Bjøn, Shengdong Zhao, Briane Paul Samson, and Rafal Kocielnik, editors, *CHI '20: CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, April 25-30, 2020*, pages 1–14. ACM, 2020.

[8] Dieter Grasbon and Norbert Braun. A morphological approach to interactive storytelling. In *Proc. CAST01, Living in Mixed Realities. Special issue of Netzspannung. org/journal, the Magazine for Media Production and Inter-media*

Research, pages 337–340. Citeseer, 2001.

[9] Erik E Guzik, Christian Byrge, and Christian Gilde. The originality of machines: Ai takes the torrance test. *Journal of Creativity*, 33(3):100065, 2023.

[10] Jennifer Haase and Paul H. P. Hanel. Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. *CoRR*, abs/2303.12003, 2023.

[11] Yingqing He, Menghan Xia, Haoxin Chen, Xiaodong Cun, Yuan Gong, Jinbo Xing, Yong Zhang, Xintao Wang, Chao Weng, Ying Shan, et al. Animate-a-story: Storytelling with retrieval-augmented video generation. *arXiv preprint arXiv:2307.06940*, 2023.

[12] Kent F Hubert, Kim N Awa, and Darya L Zabelina. The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Scientific Reports*, 14(1):3440, 2024.

[13] Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Madry, Alex Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, Alex Nichol, Alex Paino, Alex Renzin, Alex Tachard Passos, Alexander Kirillov, Alexi Christakis, Alexis Conneau, Ali Kamali, Allan Jabri, Allison Moyer, Allison Tam, Amadou Crookes, Amin Tootoonchian, Ananya Kumar, Andrea Vallone, Andrej Karpathy, Andrew Braunstein, Andrew Cann, Andrew Codispoti, Andrew Galu, Andrew Kondrich, Andrew Tulloch, Andrey Mishchenko, Angela Baek, Angela Jiang, Antoine Pelisse, Antonia Woodford, Anuj Gosalia, Arka Dhar, Ashley Pantuliano, Avi Nayak, Avital Oliver, Barret Zoph, Behrooz Ghorbani, Ben Leimberger, Ben Rossen, Ben Sokolowsky, Ben Wang, Benjamin Zweig, Beth Hoover, Blake Samic, Bob McGrew, Bobby Spero, Bogo Giertler, Bowen Cheng, Brad Lightcap, Brandon Walkin, Brendan Quinn, Brian Guarraci, Brian Hsu, Bright Kellogg, Brydon Eastman, Camillo Lugaresi, Carroll L. Wainwright, Cary Bassin, Cary Hudson, Casey Chu, Chad Nelson, Chak Li, Chan Jun Shern, Channing Conger, Charlotte Barette, Chelsea Voss, Chen Ding, Cheng Lu, Chong Zhang, Chris Beaumont, Chris Hallacy, Chris Koch, Christian Gibson, Christina Kim, Christine Choi, Christine McLeavey, Christopher Hesse, Claudia Fischer, Clemens Winter, Coley Czarnecki, Colin Jarvis, Colin Wei, Constantin Koumouzelis, and Dane Sherburn. Gpt-4o system card. *CoRR*, abs/2410.21276, 2024.

[14] Nanna Inie, Jeanette Falk, and Steve Tanimoto. Designing participatory AI: creative professionals' worries and expectations about generative AI. In Albrecht Schmidt, Kaisa Väänänen, Tesh Goyal, Per Ola Kristensson, and Anicia Peters, editors, *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, CHI EA 2023, Hamburg, Germany, April 23-28, 2023*, pages 82:1–82:8. ACM, 2023.

[15] Felix Kistler and Elisabeth André. User-defined body gestures for an interactive storytelling scenario. In *Human-Computer Interaction–INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part II 14*, pages 264–281. Springer, 2013.

[16] Luís Leite and Veronica Orvalho. Shape your body: control a virtual silhouette using body motion. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 1913–1918. 2012.

[17] Hui Liang, Jian Chang, Ismail Khalid Kazmi, Jian-Jun Zhang, and Peifeng Jiao. Puppet narrator: Utilizing motion sensing technology in storytelling for young children. In Per Backlund, Henrik Engström, and Fotis Liarokapis, editors, *7th International Conference on Games and Virtual Worlds for Serious Applications, VS-Games 2015, Skövde, Sweden, September 16-18, 2015*, pages 1–8. IEEE Computer Society, 2015.

[18] Jing Lin, Ailing Zeng, Shunlin Lu, Yuanhao Cai, Ruimao Zhang, Haoqian Wang, and Lei Zhang. Motion-x: A large-scale 3d expressive whole-body human motion dataset. *Advances in Neural Information Processing Systems*, 2023.

[19] Sang-Na Lin, Meng-Yang Chen, Na Chen, and Min-Zhi Shao. Interactive cartoon animation through collaborative storytelling: A field study. In *Proceedings of the 14th Conference on Creativity and Cognition*, pages 505–509, 2022.

[20] Federico Magni, Jiyoung Park, and Melody Manchi Chao. Humans as creativity gatekeepers: Are we biased against ai creativity? *Journal of Business and Psychology*, 39(3):643 – 656, 2024-06.

[21] Kory W. Mathewson and Piotr Mirowski. Improvised theatre alongside artificial intelligences. In Brian Magerko and Jonathan P. Rowe, editors, *Proceedings of the Thirteenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE-17), October 5-9, 2017, Snowbird, Little Cottonwood Canyon, Utah, USA*, pages 66–72. AAAI Press, 2017.

[22] Kory W. Mathewson and Piotr Mirowski. Improbotics: Exploring the imitation game using machine intelligence in improvised theatre. In Jonathan P. Rowe and Gillian Smith, editors, *Proceedings of the Fourteenth AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE 2018, November 13-17, 2018, Edmonton, Canada*, pages 59–66. AAAI Press, 2018.

[23] MediaPipe. Gesture recognition task guide, 2025.

[24] MediaPipe. Pose landmark detection guide, 2025.

[25] Piotr Mirowski, Kory W. Mathewson, Boyd Branch, Thomas Winters, Ben Verhoeven, and Jenny Elfving. Rosetta code: Improv in any language. In F. Amílcar Cardoso, Penousal Machado, Tony Veale, and João Miguel Cunha, editors, *Proceedings of the Eleventh International Conference on Computational Creativity, ICCC 2020, Coimbra, Portugal, September 7-11, 2020*, pages

115–122. Association for Computational Creativity (ACC), 2020.

[26] Piotr Mirowski and Kory Wallace Mathewson. Human improvised theatre augmented with artificial intelligence. In Steven Dow, Mary Lou Maher, Andruid Kerne, and Celine Latulipe, editors, *Proceedings of the 2019 ACM SIGCHI Conference on Creativity and Cognition, C&C 2019, San Diego, CA, USA, June 23-26, 2019*, pages 527–530. ACM, 2019.

[27] Piotr W. Mirowski, Boyd Branch, and Kory Wallace Mathewson. The theater stage as laboratory: Review of real-time comedy LLM systems for live performance. *CoRR*, abs/2501.08474, 2025.

[28] Tomáš Musil, Klára Vosecká, and Rudolf Rosa. Performing ai-generated theater plays. In *Choreomata*, pages 390–400. Chapman and Hall/CRC, 2023.

[29] OpenAI. Introducing whisper, 2022.

[30] OpenAI. Introducing the realtime api, 2024.

[31] OptiTrack. Optitrack documentation, 2025.

[32] Zhenhui Peng, Xingbo Wang, Qiushi Han, Junkai Zhu, Xiaojuan Ma, and Huamin Qu. Storyfier: Exploring vocabulary learning support with text generation models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–16, 2023.

[33] Zhen Ren. The role of ai in theatre: Exploring the creation of ai-generated stage plays. *Theoretical and Natural Science*, 34:27–32, 2024.

[34] Rudolf Rosa, Ondrej Dusek, Tom Kocmi, David Marecek, Tomás Musil, Patrícia Schmidtová, Dominik Jurko, Ondrej Bojar, Daniel Hrbek, David Kosták, Martina Kinská, Josef Dolezal, and Klára Vosecká. Theaitre: Artificial intelligence to write a theatre play. In Alípio Mário Jorge, Ricardo Campos, Adam Jatowt, and Akiko Aizawa, editors, *Proceedings of AI4Narratives - Workshop on Artificial Intelligence for Narratives in conjunction with the 29th International Joint Conference on Artificial Intelligence and the 17th Pacific Rim International Conference on Artificial Intelligence (IJCAI 2020), Yokohama, Japan, January 7th and 8th, 2021 (online event due to Covid-19 outbreak)*, volume 2794 of *CEUR Workshop Proceedings*, pages 9–13. CEUR-WS.org, 2020.

[35] Karl Toby Rosenberg, Rubaiat Habib Kazi, Li-Yi Wei, Haijun Xia, and Ken Perlin. Drawtalking: Building interactive worlds by sketching and speaking. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, pages 1–25, 2024.

[36] Jose Ma Santiago III, Richard Lance Parayno, Jordan Aiko Deja, and Briane Paul V Samson. Rolling the dice: Imagining generative ai as a dungeons & dragons storytelling companion. *arXiv preprint arXiv:2304.01860*, 2023.

[37] Hanieh Shakeri, Carman Neustaedter, and Steve DiPaola. SAGA: collaborative storytelling with GPT-3. In Jeremy P. Birnholtz, Luigina Ciolfi, Sharon Ding, Susan R. Fussell, Andrés Monroy-Hernández, Sean Munson, Irina Shklovski, and Mor Naaman, editors, *Companion Publication of the 2021 ACM Conference on Computer Supported Cooperative Work and Social Computing,*

CSCW 2021, Virtual Event, USA, October 23-27, 2021, pages 163–166. ACM, 2021.

[38] David Thue, Vadim Bulitko, Marcia Spetch, and Eric Wasylishen. Interactive storytelling: A player modelling approach. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 3, pages 43–48, 2007.

[39] Florent Vinchon, Todd Lubart, Sabrina Bartolotta, Valentin Gironnay, Marion Botella, Samira Bourgeois-Bougrine, Jean-Marie Burkhardt, Nathalie Bonnardel, Giovanni Emanuele Corazza, Vlad Glăveanu, Michael Hanchett Hanson, Zorana Ivcevic, Maciej Karwowski, James C. Kaufman, Takeshi Okada, Roni Reiter-Palmon, and Andrea Gaggioli. Artificial intelligence & creativity: A manifesto for collaboration. *The Journal of Creative Behavior*, 57(4):472–484, 2023.

[40] Zhuohao Wu, Danwen Ji, Kaiwen Yu, Xianxu Zeng, Dingming Wu, and Mohammad Shidujaman. AI creativity and the human-ai co-creation model. In Masaaki Kurosu, editor, *Human-Computer Interaction. Theory, Methods and Tools - Thematic Area, HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24-29, 2021, Proceedings, Part I*, volume 12762 of *Lecture Notes in Computer Science*, pages 171–190. Springer, 2021.

[41] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. Wordcraft: story writing with large language models. In *Proceedings of the 27th International Conference on Intelligent User Interfaces*, pages 841–852, 2022.

[42] Chao Zhang, Cheng Yao, Jiayi Wu, Weijia Lin, Lijuan Liu, Ge Yan, and Fangtian Ying. Storydrawer: a child–ai collaborative drawing system to support children's creative visual storytelling. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, pages 1–15, 2022.

[43] Lei Zhang, Daekun Kim, Youjean Cho, Ava Robinson, Yu Jiang Tham, Rajan Vaish, and Andrés Monroy-Hernández. Jigsaw: Authoring immersive storytelling experiences with augmented reality and internet of things. In Florian 'Floyd' Mueller, Penny Kyburz, Julie R. Williamson, Corina Sas, Max L. Wilson, Phoebe O. Toups Dugas, and Irina Shklovski, editors, *Proceedings of the CHI Conference on Human Factors in Computing Systems, CHI 2024, Honolulu, HI, USA, May 11-16, 2024*, pages 87:1–87:14. ACM, 2024.

[44] Zheng Zhang, Ying Xu, Yanhao Wang, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, and Toby Jia-Jun Li. Storybuddy: A human-ai collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–21, 2022.

Acknowledgements

Giunto al termine di questo elaborato e soprattutto di questo percorso, voglio dedicare qualche parola a chi mi è stato vicino in questi anni pieni di studio, racconti e tanti bei momenti. Il percorso è stato lungo e faticoso, ma indimenticabile grazie a tutti voi.

Ringrazio il professor Sanna, per avermi aiutato nella stesura di questa tesi e per essere sempre stato a disposizione, con le sue risposte celeri e dettagliate. Il suo aiuto è stato fondamentale per portare a termine questo percorso nonostante tutte le difficoltà incontrate.

Voglio ringraziare tutti quelli con cui ho condiviso parte della mia strada, ma anche quelli incontrati per pochi passi.

Ringrazio Gabbo e Michi per tutte le risate, l'aiuto e i consigli durante tutti i momenti di studio e di svago passati insieme.

Ringrazio Ari per essere stata un'amica sincera, con la quale ho condiviso tanto anche a distanza.

Ringrazio Brigi per le risate, la disperazione e il suo affetto. È stato bello scoprire il tuo grande cuore.

Ringrazio Letizia per i suoi abbracci (unici, non c'è paragone), e per essere un'amica premurosa e speciale.

Ringrazio Matteo, collega, "compagno di squadra" e di serate.

Ringrazio il mio gruppo di triennale, per avermi aiutato ad ambientarmi, e tutti gli amici della magistrale, per avermi accompagnato in tanti cambiamenti. Un grazie va a tutti i colleghi che mi sono stati vicini.

Un grazie speciale va a Vlad, mio compagno di università, di serate e fan di Piazza Statuto. Ti ringrazio per essermi stato vicino, per avermi condiviso parte della tua vita e per ogni singolo messaggio (anche quelli a orari improponibili). Da Torino all'Abruzzo, passando per il Portogallo, la Sicilia, Milano e sicuramente anche l'Inghilterra, sono sicuro che viaggeremo per aggiungere racconti al nostro libro (e per aggiungere viaggi alla tua applicazione).

Avrei tante cose da dire per Emanuele, amico trovato per caso ma che non si scollerà facilmente. Sin dalla mia prima visita a Pescara, il nostro viaggio è stato pieno di sorprese, di battute, di disperazione e anche complicità. Grazie per avermi offerto dimora nelle serate più disparate e per essermi stato vicino anche a 10.000 Km di distanza. Non dimenticherò mai le serie guardate insieme, le tue doti con la pallina da ping pong, i 10 di quel mazzo di carte, il tuo dente – mi rimarrà sempre in testa – e anche il buonissimo tavernello. Non posso citare tutte le storie, ma fatti bastare un "grazie per tutto" (palese hai fatto peggio nella tua tesi, sono più bravo io a scrivere cose carine).

Ringrazio il collegio e tutti gli amici che ho conosciuto tramite questo posto mistico e molto discutibile. Voglio ringraziare MatBat il videomaker, FrancescoCagn il fenomeno, Juan l'uomo budino, Cucci il mio maestro dell'informatica, Alessandro – Muzzi – e Mario. Vi ringrazio per tutti i bei momenti passati insieme e il supporto che mi avete dato.

Ringrazio Stefano, mio compagno di palestra (anche se per poco), consulente fidato e con un buon gusto sugli anime ma non sul cibo. Ti ringrazio per avermi fatto divertire e soprattutto per avermi fatto riflettere.

Non bastano le parole per descrivere il rapporto con "gli amici di Tanino". Grazie Aurelio per avermi fatto compagnia nel bullizzare il tuo gemello, per averci fatto ridere e per essere un caro collega siciliano, condividendo conoscenze e riferimenti inusuali. Grazie Dadda (Gaetano, Davide o anche Topolino) per i tuoi sorrisi, le tue vocine e tutte le serate passate insieme – ti riporterò a casa tutte le volte che servirà. Grazie Francesco, ti ringrazio per avermi e averci fatto da mamma/papà/zio, grazie per averci salvato da 7 piadine alla settimana, grazie per i tuoi racconti e la tua infinità disponibilità. Infine, ringrazio Vincenzo per essere stato un grande amico, sempre pronto a offrire un consiglio o anche soltanto un supporto, per un lamento o per discutere del più incredibile avvenimento. Grazie capitano per avermi fornito tanti assist, sul campo e nella vita. Mi mancano le crepes del lunedì sera.

Senza dimenticarmi da dove provengo, un pensiero va a tutti i miei amici calatini. Ringrazio Ciccio, Franco, Carlo, Simone, Manu, Matteo, per avermi reso lieti i momenti di vacanza, aggiungendo aneddoti per i futuri noi. Ringrazio il caro Giuseppe e Criscio per aver portato un po' di Caltagirone e un po' di Libertas a Torino.

Ringrazio Asia e Tecla, per i racconti e i tanti ricordi, ma soprattutto per le risate insieme.

Rigrazio Peppe per essere stato un amico fedele, sempre a disposizione e pronto ad ascoltare i miei lunghi audio, accompagnandomi nelle lunghe passeggiate notturne.

Fra i calatini-torinesi, bisogna citare le mie superchicche, Nara e Giuliana, che mi hanno fatto compagnia prima nella sezione G, poi nella A, e infine a Torino. Ci siamo ritrovati in un mondo più grande di noi, ma siete sempre state un supporto e una piccola famiglia. Grazie per avermi fatto sentire amato e per aver gioito a ogni mio traguardo.

Ringrazio anche una costante della mia vita, a volte per un messaggio, a volte per una battuta. Dal liceo hai fatto parte della mia quotidianità, dicendo cose relativamente intelligenti ma sempre con un grande affetto nei miei confronti. Grazie Sofia.

A special thanks or rather an "arigatou gozaimashita" goes to all the members of the sinking ship, Simon, Ioana, Matteo, Suraj, Mus, Igor.

Thanks to Ivan, for being my first support in this bizarre adventure. You were a great traveling companion, especially for your absurd speeches.

Thanks to Claire for making me laugh for your questionable choices and for the continuous updates on your life.

Thanks Harald for all the absurd nights with Japanese strangers (a thought goes to Satou).

Grazie Adil per essere stato un grande amico. Nonostante tu rimanga molto e forse troppo cinico, ogni tanto si vede la parte buona di te. Sei una persona affettuosa che tiene alle sue amicizie, mostrandolo con piccoli gesti. Ti ringrazio per avermi aiutato nella stesura di questa tesi, per avermi accolto nella tua famiglia e per avermi trattato come un fratello.

Ringrazio Oriana per la risata, l'accento napoletano e tutto l'affetto.

Due parole vanno spese per una persona entrata da poco nella mia vita, ma che ha già lasciato il segno. Grazie Gaia per tutto il tempo passato insieme, per ogni singola chiamata, messaggio vocale, cena al Mc e ogni ricordo condiviso – soprattutto la tua performance su YouTube. Grazie per avermi fatto sentire amato, per avermi supportato e per avermi fatto ridere. Mi sento molto fortunato ad averti conosciuto e sono grato per ogni tuo singolo sorriso.

Il mio ringraziamento più grande va ai miei parenti, ai miei zii, Antonio e la nonna Italia.

Ringrazio nonna Filippa, lei che mi hai cresciuto con tanto l'amore, più di quanto io possa descrivere a parole. Grazie per ogni singolo momento insieme.

Ringrazio Marta, mia madrina, mia guida e mia compare. Sin da piccoli ti ho sempre preso come esempio e penso a te anche adesso, nel momento in cui il mio percorso si conclude. Sarai sempre un punto di riferimento non solo nello studio ma nella vita.

Infine, ringrazio la mia squadra, il mio rifugio, i miei consiglieri, il mio fan club, la mia famiglia.

Ringrazio mamma e papà, che mi pensano e si preoccupano sempre per i figli. Siete sempre premurosi e disposti a fare qualsiasi sacrificio per renderci felici, senza farci mancare mai niente. Mi avete sempre supportato nelle mie scelte ed è proprio grazie a voi che ho raggiunto tutti questi traguardi. Grazie per tutto il vostro amore.

Ringrazio Giulia, la mia fonte di sticker preferita e non solo. Ti ringrazio per tutto l'aiuto e l'idee di questi mesi. Sei la mia compagna di vita, con cui condividere tutti i litigi dell'infanzia ma anche tutti i consigli della maturità. Sei una persona speciale, attenta e soprattutto intelligente (solo emotivamente – ti ho visto piangere dal ridere per video troppo discutibili). È sempre bello ridere con te. Ti voglio tanto bene.

Grazie a voi per avermi guidato e per aver festeggiato i miei traguardi.

Ringrazio Diego per la compagnia in giornate solitarie.