



Patrick Struger, BSc

# **Learning with AI based videos**

## **Master's Thesis**

to achieve the university degree of

Master of Science

Master's degree programme: Computer Science

submitted to

**Graz University of Technology**

Supervisor

Priv.-Doz. Dipl.-Ing. Dr.techn. Martin Ebner,  
Benedikt Brünner, MEd BEd

Institute of Interactive Systems and Data Science

Head: Univ.-Prof. Dipl.-Ing. Dr.techn. Frank Kappe

Graz, December 2024

This document is set in Palatino, compiled with [pdfL<sup>A</sup>T<sub>E</sub>X2e](#) and [Biber](#).

The L<sup>A</sup>T<sub>E</sub>X template from Karl Voit is based on [KOMA script](#) and can be found online: <https://github.com/novoid/LaTeX-KOMA-template>

## Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

---

Date

---

Signature

# Abstract

The progressive development of Artificial Intelligence (AI) driven applications and learning platforms for knowledge transfer and content generation in the field of education is giving rise to new research topics that focus on the influence on numerous aspects of learning. However, the development in the field of AI has been driven forward in many areas for a long time and ranges from recommender systems to the analysis of video materials a few years back. Due to the constant further development, especially in the area of content creation, the education and learning sectors are now also being more widely perceived and focused on by the public. In the course of this thesis, the use of learning videos, created with AI human avatars, is examined in more detail and evaluated on the basis of emotional response and user feedback. The aim of this thesis is to identify significant differences between learning videos showing artificially generated human avatars and real lecturers in terms of cognitive and emotional engagement of learners and to draw possible conclusions about the perception and efficiency of current applications of generative AI in learning environments. For this purpose the empirical study of this thesis is based on a qualitative and quantitative analysis of 55 participants from different academic fields, whose learning experience and learning outcome were recorded and evaluated using self-produced learning videos with artificially generated and real lecturers with post assessments and facial analysis software. The personal assessment and perception of the participants was then additionally recorded through face-to-face interviews and subjected to a qualitative content analysis. The thesis is divided into an introductory part to provide the necessary background knowledge, followed by the underlying research method which was used for the framework and implementation of the study, the result part in which the aspects to be analyzed and their statistical outcome are described and classified, and finally at the end of the thesis a more detailed discussion and respective conclusions are drawn.

# Contents

<b>Abstract</b>	<b>iv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Learning Environments . . . . .	3
1.1.1 Learning Management Systems . . . . .	3
1.1.2 Intelligent Learning Management Systems . . . . .	4
1.1.3 MOOC . . . . .	5
1.2 Learning Content Delivery . . . . .	6
1.2.1 Educational Videos . . . . .	6
1.2.2 Instructor Presence . . . . .	8
1.2.3 Learning videos with Artificial Intelligence . . . . .	12
1.3 Learning Engagement . . . . .	14
1.3.1 Behavioral Learning Engagement . . . . .	14
1.3.2 Emotional Learning Engagement . . . . .	15
1.3.3 Cognitive Learning Engagement . . . . .	15
1.3.4 Measurement of Learning Engagement . . . . .	16
1.3.5 Digital Inclusion . . . . .	17
<b>2 Research Method</b>	<b>18</b>
2.1 Participants . . . . .	18
2.2 Video Creation . . . . .	19
2.3 Experiment Setup . . . . .	21
2.4 Implementation . . . . .	23
2.5 Course Design . . . . .	24
2.6 Evaluation . . . . .	25
2.6.1 Facial Analysis . . . . .	25
2.6.2 Post Assessment . . . . .	25
2.6.3 Problem-centered Interview . . . . .	25

## Contents

<b>3</b>	<b>Results</b>	<b>32</b>
3.1	Post Assessments . . . . .	32
3.2	Emotional Classification . . . . .	35
3.3	Problem Centered Interview . . . . .	44
3.4	Stimulus Video . . . . .	62
<b>4</b>	<b>Discussion</b>	<b>66</b>
<b>5</b>	<b>Conclusion</b>	<b>69</b>
<b>6</b>	<b>Outline</b>	<b>70</b>
	<b>Bibliography</b>	<b>71</b>
	<b>Appendix</b>	<b>75</b>
A.1	Consent form for participants . . . . .	76
A.2	Interview Questions . . . . .	78
A.3	Post Assessment Questions . . . . .	79
A.4	FaceReader Online Result View . . . . .	80

# List of Figures

1.1	LMS diagram, adapted from Van Vaerenbergh and Pérez-Suay, 2022 . . . . .	3
1.2	Intelligent LMS diagram, adapted from Van Vaerenbergh and Pérez-Suay, 2022 . . . . .	4
1.3	CASA Paradigm of Educational Videos as Social Actor, adapted from CASA paradigm by Nass, Steuer, and Siminoff, 1994 . .	9
2.1	Video creation using a Panasonic GH5s camera with Olympus 12-200mm lens with 4k, 25fps, ISO 400, 1/50 shutter speed and 5.6 aperture considering 2m distance to the well-lit green screen and 2m to the front camera. . . . .	19
2.2	Video creation using the AI-powered video tool HeyGen provided with footage of the real instructor. . . . .	20
2.3	From left to right: Video frame with the real and AI-generated presenter of the male instructor used for the study, created in Adobe Premiere Pro. . . . .	21
2.4	Experiment setting at the educational technologies organizational unit of TU Graz. . . . .	22
2.5	Flow chart of the process of study implementation. . . . .	23
2.6	Structure of micro learning courses designed for this study. .	24
2.7	Process model of inductive category formation adapted from Mayring and Fenzl, 2019. . . . .	28
3.1	Discrete distribution of wrong and right answers to post assessment of learning videos with real instructor and virtual instructor. . . . .	33
3.2	Participants' answers to post assessments grouped by the two course topics "iMooX" and "MetaCampus". . . . .	34

## List of Figures

3.3	Comparison of captured emotional states using FaceReader software of learning videos with real instructors (top) and virtual instructors (bottom) in percent. . . . .	36
3.4	Quality ranking of the facial analysis software regarding recordings of the participants. . . . .	37
3.5	Comparison of classified emotions between real and virtual male instructor of the first learning course. . . . .	38
3.6	Comparison of classified emotions between real and virtual female instructor of the first learning course. . . . .	38
3.7	Comparison of classified emotions between real and virtual male instructor of the second learning course. . . . .	39
3.8	Comparison of classified emotions between real and virtual female instructor of the second learning course. . . . .	39
3.9	Differences between valence, arousal, responsiveness and attention during the micro-courses comparing virtual and real instructor. . . . .	41
3.10	The Arousal-Valence Model of Emotions adapted from Elamir, Al-Atabany, and Eldosoky (2019). . . . .	42
3.11	The Arousal-Valence Model of (N=55) participants. . . . .	43
3.12	Distribution of the amount of participants recognized AI generated instructors. . . . .	46
3.13	Number of participants noticing AI instructors grouped by prior knowledge of the corresponding real instructor. . . . .	47
3.14	Preference of participants using AI generated instructors in learning videos. . . . .	54
3.15	Participants who would generate an AI avatar of themselves grouped by their educational background. . . . .	55
3.16	Participants who would generate an AI avatar of themselves grouped by their age. . . . .	56
3.17	AI Avatar preferences of participants . . . . .	57
3.18	Comparison of the real female and virtual instructor at the end of the learning session. . . . .	62
3.19	Participants emotional journey during the final stimulus video comparing the real male and the virtual male instructor with the strongest basic emotional expressions. . . . .	63



## List of Figures

3.20	Participants emotional journey during the final stimulus video comparing the real female and the virtual female instructor with the strongest basic emotional expressions. . . .	65
.1	First page of the consent form for participants for the recording and evaluation of biometric data with the FaceReader software as well as their legally compliant storage, processing and deletion. . . . .	76
.2	Second page of the consent form for participants for the recording and evaluation of biometric data with the FaceReader software as well as their legally compliant storage, processing and deletion. . . . .	77
.3	Interview questions of the problem centered interview consisting of 3 topic blocks with 1 key question and corresponding sub-questions. . . . .	78
.4	Interactive result view for participant #4 concerning the micro-course with the topic "MetaCampus" provided through the analytical backend of FaceReader Online. . . . .	80

# 1 Introduction

Research regarding the evolutionary process of learning supported by educational technology and technical approaches like e-learning or distance learning has been carried out since 1970. Especially the last 20 years revealed an exponential increase of publications in this area while many of them have focused on specific applications such as gamification, virtual reality, artificial intelligence, mobile learning and augmented reality according to Martinez-Garcia, Horrach-Rosselló, and Mulet-Forteza (2023).

Significant findings, definitions and achievements in the technical and social domain based on those researches have revolutionized learning for future generations permanently. Areas like Social and Emotional Learning (SEL) play an important role and are going to be increasingly integrated in Learning Management Systems (LMS).

Due to technical achievements and possibilities in management, documentation, tracking and automation in Learning Management Systems, delivery of educational content is increasingly shifting into the digital world. Learning videos are traded as the main source of educational content delivery in such systems while producing high quality content is being more and more supported by techniques of the artificial domain. Use-cases like transcription, translation, text generation or even generation of artificial avatars with emotional bonding techniques are being increasingly integrated in state-of-the-art Learning Management Systems.

Although educational content can easily be produced and delivered by Artificial Intelligence (AI), only a few studies like Leiker et al. (2023) or Pataranutaporn et al. (2022) cover the overall learning outcome and emotional aspects of students in higher education when using such techniques

## 1 Introduction

in learning environments. Thus, this study aims to answer the following Research Questions (RQ):

- **RQ1:** Does the use of AI avatars in learning videos affect their quality from the learners' perspective?
- **RQ2:** Is the recognition software FaceReader Online<sup>1</sup> able to track reliable emotional states while watching learning videos?
- **RQ3:** Is there a significant difference in emotion and therefore in Emotional Learning when using AI generated presenters?

**To answer those questions understandable and in an appropriate manner, the whole thesis follows the IMRaD<sup>2</sup> structure.**

The introductory chapter with section 1.1, section 1.2 and section 1.3 of this thesis focuses on relevant background knowledge covering and summarizing core concepts, examples and current methodology based on educational content delivery and the impact of learning engagement in learning management systems in combination with artificial technology.

Chapter 2 describe the conducted case study, its implementation as well as the framework of the underlying research method focusing on the impact of educational video content presented by virtual avatars concerning the perception and emotional behavior in such environment settings. Furthermore used tools, equipment and software for the study are being described in more detail. Results are being presented in chapter 3 giving useful insights into students' learning experience and emotional participation in environments delivering educational content with the help of generative AI in higher education. Chapter 4 provides answers to prior defined research questions derived by analyzing the received results. Finally the thesis is concluded in chapter 5 and gives an outline for future research and considerations in chapter 6.

---

<sup>1</sup><https://facereader-online.com/> (last access November 2024)

<sup>2</sup><https://en.wikipedia.org/wiki/IMRAD> (last access September 2024)

## 1.1 Learning Environments

### 1.1.1 Learning Management Systems

A Learning Management System (LMS) describes a virtual platform to create, deliver and report on training courses as well as development and learning programs pointed out by Van Vaerenbergh and Pérez-Suay, 2022. The usage of LMS has grown during the last decades continuously, especially in higher education. Most LMS offer interactive learning environments, allowing students to access content independent of location and time, to collaborate and communicate with teachers, peers and experts. Prominent examples of such LMS are Moodle<sup>3</sup>, Blackboard<sup>4</sup> or Coursera<sup>5</sup>. The main parts of standard functionalities provided by current LMS are illustrated in Figure 1.1

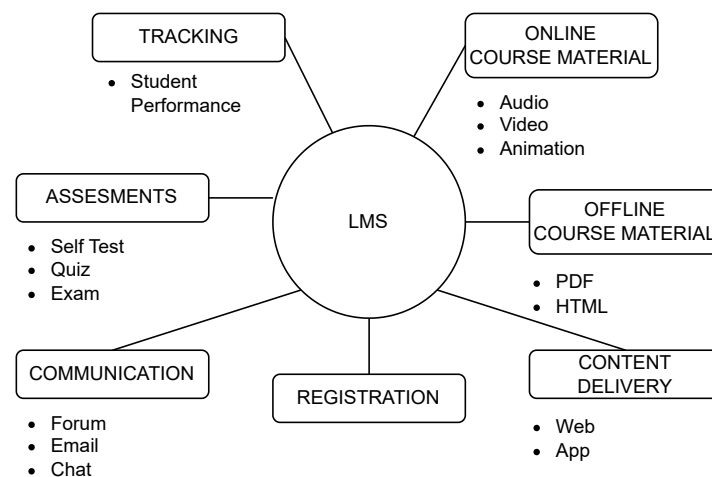


Figure 1.1: LMS diagram, adapted from Van Vaerenbergh and Pérez-Suay, 2022

<sup>3</sup><https://moodle.com/> (last access September 2024)

<sup>4</sup><https://www.blackboard.com/> (last access September 2024)

<sup>5</sup><https://www.coursera.org/> (last access September 2024)

### 1.1.2 Intelligent Learning Management Systems

The concept of Intelligent Learning Management Systems (ILMS) extends traditional LMS functionality with intelligent features like predictive modeling, interactions and automated tasks provided by AI. Since the field of AI is very broad, the definition of an ILMS in the educational context can be derived by extending ordinary LMS functionalities by the following two aspects according to Van Vaerenbergh and Pérez-Suay (2022):

- **Learning Analytics**

Through continuously tracking and acquiring data by the LMS, learning analytics techniques can be applied to model student data and extract useful insights on learning progress and behaviour. Analytical results offer the opportunity of *Adaptive Learning Systems (ALS)* where the behaviour of the LMS itself can be adapted through interfaces, course flow, organization or difficulty to specific student models accordingly.

- **Intelligent Tutoring**

Through classification and extraction of written text, spoken text, drawings or user input the system is capable of personalized real-time feedback or providing answers to specific questions (e.g. Chatbots).

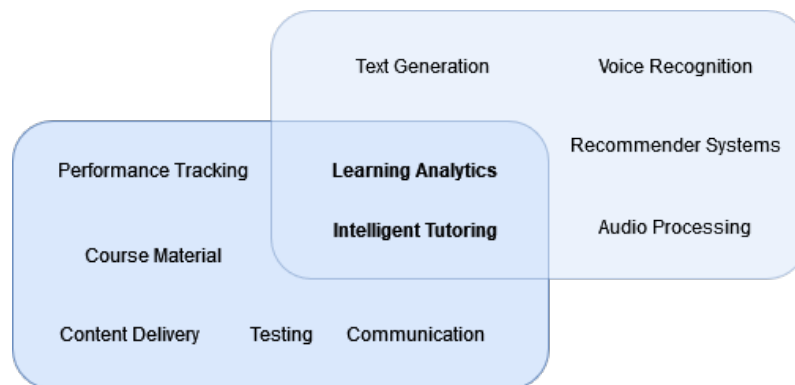


Figure 1.2: Intelligent LMS diagram, adapted from Van Vaerenbergh and Pérez-Suay, 2022

### 1.1.3 MOOC

The term MOOC stands for *Massive Open Online Course* and describes a specific type of course available with no admission restrictions and free of charge. MOOCs are used for content delivery and managing learning experience for a broad audience. Since the first MOOC was held by Stephen Downes, 2008 different formats have emerged through the last years:

- **cMOOC**  
Courses are designed in a connectivist or constructivist manner, i.e. the seminar-like course form originally intended by Stephen Downes, 2008 without clearly defined learning objectives focusing on whole networks of participants.
- **xMOOC**  
Courses are based primarily on video content and typically provide simple assessment tools for defined learning objectives such as quizzes and communication opportunities such as discussion forums focusing on the individual participant.
- **bMOOC**  
Courses are designed in a blended setting with seminar-like face-to-face events combined with online learning components. The bMOOC format can be seen as the convergence of cMOOC, xMOOC, and face-to-face learning.
- **smOOC**  
Courses are designed for smaller groups, e.g. 100-150 participants and exhibit significant impact on social inclusion processes according to Marta-Lazo, Osuna-Acedo, and Gil-Quintana (2019).

The impact of MOOCs in learning environments especially on analyzing students behavioral engagement and motivation as predictors of learning outcomes has become a prominent research topic over the last few years. Strong shifts towards home schooling and distance learning during the corona pandemic accompanied by a significant increase in digitization in general, caused various research topics on online content delivery.

### 1.2 Learning Content Delivery

The manner of learning content delivery in LMS can be done in many different ways and indicates a strong impact on students' learning experience. Depending on the learning objective, educational content can be communicated via audio and video files, animations, course slides or text-based.

#### 1.2.1 Educational Videos

Educational video content has become an important part in higher education specifically and is explored as a highly effective educational tool shown by several meta-analysis and studies summarized by Brame (2016). Learning videos are integrated as part of traditional courses, serve as a cornerstone of many blended courses, and are often the main information-delivery mechanism in online courses which have become a prominent choice for home schooling or distance learning conveying learning content to many people with absence of physical presence.

According to Brame (2016) effective development or use of videos as an educational tool is enhanced when considering students' cognitive load, engagement and enabling active learning.

#### Cognitive Load

Building on the theory, that working memory has two channels for recording and processing information. A visual/pictorial channel and an auditory/verbal-processing channel. Effective implementations of considering cognitive load in multimedia learning settings recommended by Brame (2016) can be:

- Use Signaling to highlight important information
- Use Segmenting to chunk information
- Use Weeding of extraneous information

## 1 Introduction

### Student Engagement

The basic idea is to get students to watch the videos otherwise the learning content won't be received and no learning success can be achieved. Several studies like Guo, Kim, and Rubin (2014) with MOOC videos show that video length and type of conversation have a strong impact on students' attention. They observed that videos less than 6 minutes long result in the longest engagement time, close to 100% of the overall video length. Videos with the length of 9-12 minutes result in 50% and 12-40 minutes in 20% of the whole video content considering students' attention. According to Mayer (2008) a quick and enthusiastic way of speaking encourages students to develop a sense of social partnership with the presenter which leads to greater engagement. Recommendations for increasing students' engagement in learning videos can therefore be derived by following two main criteria:

- Keep videos brief and targeted on learning goals
- Use natural and enthusiastic conversational styles

### Active Learning

In the context of online learning, e.g. via MOOCs, watching videos can lead to a more passive way of consuming educational content in general. Schacter and Szpunar (2015) identified online learning as a type of self-regulated learning where the promotion of cognitive activity during watching learning videos is highly important for enhancing learning from educational videos. Depending on the learning objective following strategies have demonstrated success during viewing educational video content:

- Interactive Questions
- Interactive Control Features
- Guiding Questions
- Videos as part of Assignments



### 1.2.2 Instructor Presence

Instructors represent a central role in communication of learning content. Whether knowledge is imparted physically or digitally, students' engagement and learning experience exhibit a strong co-dependence for effective or ineffective content delivery. The presence or absence of the instructor is concerned as one of the fundamental design aspects in learning videos. There are several different approaches to include the instructor in learning videos:

- **Fully Visible**

Instructor acts in the centre of the video, communicating in a more "lecture-style" way of presenting. Students perceive social cues like direct eye contact and gestures of the instructor.

- **Next to Screen**

Instructor is visible on the left or on the right side of the video while the content is placed in a more centered way. This approach tries to get the actual content into the focus while it is commented and accompanied by the fully visible instructor by voice and gestures.

- **Picture-in-Picture**

The "Talking Head" principle uses a picture-in-picture overlaid in learning videos and presents the content throughout the whole video besides a small clip of the instructors' face commenting and explaining the currently addressed topic. This approach does not allow students to perceive gestures or body language of the presenter.

Summarized findings by Ng and Przybyłek (2021) concerning the effects of instructor presence in learning videos, state that learning outcomes, perceptions or attention allocation are rather inconsistent, strongly depend on the learning objective and can have positive, negative or even no effects on students' learning progress during watching educational video content. Effects with the most remarkable impact on influencing learning outcome can be categorized in *Social Presence*, *Gesture Effects*, *Para-social Interaction*, the *Split Attention Effect* and the *Image Principle*.

## 1 Introduction

### Social Presence

According to the Social Agency Theory by Reeves and Nass (1996) and the CASA paradigm (*Computers-Are-Social-Actors*) by Nass, Steuer, and Siminoff (1994) illustrated in Figure 1.3, video-based instruction can be interpreted as a social event. Social cues like voice, gestures or eye contact, cause a social activation scheme triggering social processes of human-to-human interaction. Thus educational videos with instructor presence are not only be interpreted as a piece of information but also as a situation of social communication. Park (2015) showed that implementing social cues in multimedia learning environments lead to lower mental load and more cognitive effort of learners. Therefore students' engagement and learning outcomes are fostered when content is delivered in a way that heightens social presence.

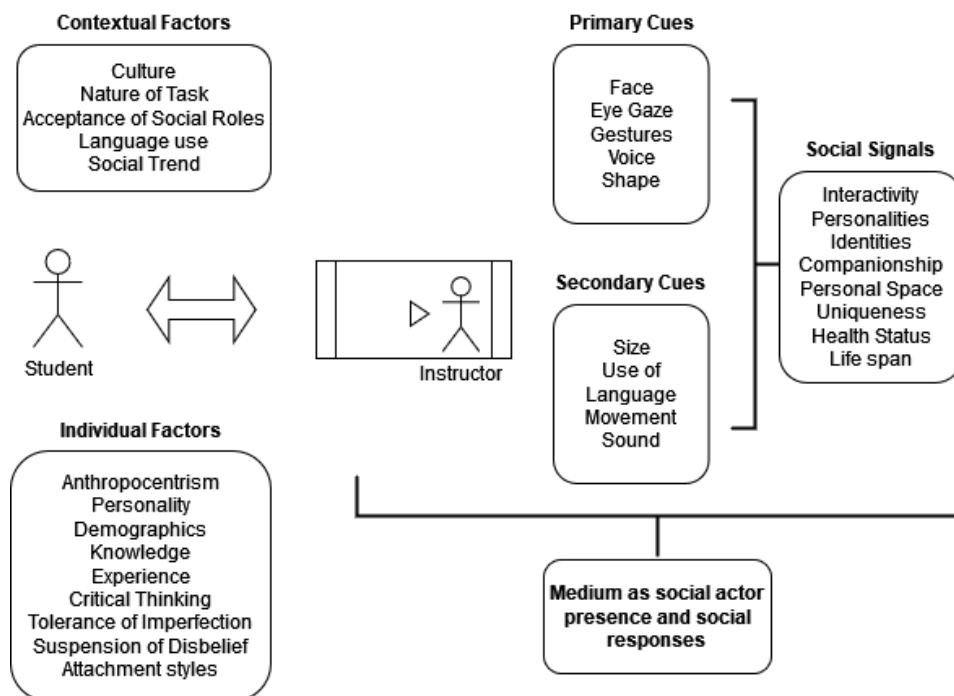


Figure 1.3: CASA Paradigm of Educational Videos as Social Actor, adapted from CASA paradigm by Nass, Steuer, and Siminoff, 1994

## 1 Introduction

### Gesture Effects

Serving as an essential aspects of human communication, instructor gestures in learning videos can be divided into beat gestures (rhythmic) and deictic gestures (signaling). According to Beege, Ninaus, et al. (2020), beat gestures support the verbal communication but also may direct the attention to the lecturer. Rhythmic gestures are characterized by concrete or abstract pointing movements directly to visual information, which are always context-sensitive. Findings by Beege, Ninaus, et al. (2020) revealed, that learning outcomes are enhanced by deictic gestures signaling visual material through increasing attention and social presence but not by beat gestures.

### Para-social Interaction

The effect of para-social interaction takes place when a social entity like persons or fictional characters are implemented in the learning environment and influence the learning process. The term "para" is attributed to the missing communication channel from the student towards the social entity, which is the case in most videos delivering educational content. Beege, Schneider, et al. (2017) shows that frontal orientation of social entities in learning videos increased the perception and retention performance in contrast to lateral presentation by giving learners the impression to be addressed directly through frontal eye contact.

### Split Attention Effect

According to cognitive load theory in multimedia learning, mentioned in 1.2.1, processing of information in the working memory has limited capacity due to finite cognitive resources. Thus in the context of learning videos, stated by Ng and Przybyłek (2021) and in contrast to the effect of social presence mentioned in section 1.2.2, the very presence of instructors carries no pedagogically relevant information but brings extraneous load to learners and occupies the limited working memory capacity which subsequently results in subpar learning, unwanted distraction or lack of concentration.

## 1 Introduction

### The Image Principle

Defined as one of the principles of Mayer (2009) derived from multimedia learning experiments and in addition to the Split Attention Effect mentioned in section 1.2.2, the image principle suggests that *"people do not necessarily learn more deeply from a multimedia presentation when the speaker's image is on the screen rather than not on the screen"* meaning that extraneous visual processing caused by the presence of the instructor would offset the potential positive effect of social presence. Although some of the findings of Ng and Przybyłek (2021) show that students who had learnt from multimedia lessons with a pedagogical agent performed slightly better but only were tested with animated agents in gamified and interactive settings.

Due to inconsistent research outcomes regarding instructor presence in learning videos, a current meta-analysis from Beege, Schroeder, et al. (2023) comprises 27 years of research regarding the instructor presence effect coming up with following highlights:

- Instructor presence was found beneficial for retention but not for transfer performance
- Social presence as well as affective and motivational ratings were positively affected by instructor presence
- In contrast to Cognitive Load Theory, instructor presence reduced cognitive load
- Instructor presence reduced dwell time on relevant visual information

### 1.2.3 Learning videos with Artificial Intelligence

Strong shifts towards online learning and blended learning and great progress of artificial intelligence over the last few years enabled various possibilities, approaches and tools to support creating and delivering educational content. Increasing demand of online learning content for degree, training and retraining programs through MOOCs, drives the need of a significant amount of instructional learning videos to enhance pedagogy and delivering the message of learning objectives.

Tools such as ChatGPT<sup>6</sup> or Dalle-2<sup>7</sup>, which are used with the more common term Generative Artificial Intelligence (GAI), have had a disruptive effect on the assessment practices in higher education concerning academic integrity, cheating and plagiarism according to Moorhouse, Yeo, and Wan (2023) but are also capable of creating media and educational content in various new ways. The main potential of using AI tools for learning videos lies in the frequently occurring conditions such as lack of time, missing on-screen experience of instructors or the lack of resources and equipment of educational institutions for producing high quality video media.

#### Content Generation

Recent statistics mentioned by Pellas (2023) show that 91% of internet users worldwide watch videos weekly and spend more than half of their daily online time engaging with general video content while the number of video streaming market share for educational content and e-learning is increasing constantly as well. The process of content creation through **Large Language Models (LLMs)**, such as GPT-4<sup>8</sup> helps in generating articles, captions, stories along with video editing and production.

---

<sup>6</sup><https://chatgpt.com> (last access November 2024)

<sup>7</sup><https://openai.com/index/dall-e-2/> (last access September 2024)

<sup>8</sup><https://openai.com/index/gpt-4/> (last access November 2024)

## 1 Introduction

Common current video platforms offering AI-generated video content, such as HeyGen<sup>9</sup> (*used for this study*), DeepBrain<sup>10</sup> or Synthesia<sup>11</sup> enable to effortlessly create videos based solely on a script within a short amount of time without requiring certain video producing skills. In the educational context, these tools can help teachers impart knowledge in an engaging way, with minimal time and cost compared to traditional video creation and can therefore be highly advantageous in online educational settings according to Leiker et al. (2023).

### Virtual Instructors

Regarding educational video content, the presenter can also be generated visually and auditory using AI with such video platforms. Either a predefined virtual avatar or a real reference model, used for imitating gestures and facial expressions, can be inserted into the learning video. This gives teachers the opportunity to impart knowledge still “personally” instead of relying on external presenters in order to continue to suggest a personal connection to the students. Since emotional learning also plays an essential role in the learning process, this thesis deals with the analysis of emotional differences between real instructors and artificially generated instructors in learning videos and examines whether this type of technology can already be used effectively and meaningfully in learning environments.

### Quality Assurance

Videos generated from scripts, deepfake videos based on face-swap technologies and animated images using reenactment technology make it harder to distinguish truthful content from fake content. Using well-considered definitions and conditions by online learning platforms and learning management systems such as proofreading of generated texts and scripts by experts, notices for students or agreeing on common guidelines and tools for content creation continue to ensure quality of educational content.

---

<sup>9</sup><https://www.heygen.com/> (last access November 2024)

<sup>10</sup><https://www.aistudios.com/de> (last access November 2024)

<sup>11</sup><https://www.synthesia.io/de> (last access November 2024)

### 1.3 Learning Engagement

When introducing new approaches and techniques in learning environments it is crucial to consider students engagement or disengagement as part of the validation process.

According to Al-Shabandar et al., 2018 students engagement can be classified in behavioral, emotional, and cognitive engagement. While behavioral engagement refers to the students participation level (e.g. number of video hits or submissions), the cognitive and emotional engagement refer to feelings regarding the progress in academic tasks or emotional engagement in learning activities.

Findings of Al-Shabandar et al., 2018 further demonstrate student's engagement in online courses for two main reasons:

- immediate satisfaction when undertaking a task
- attaining formal recognition by obtaining a certificate

#### 1.3.1 Behavioral Learning Engagement

Behavioral learning engagement refers to the observable actions and participation of students during learning activities. In the case of learning engagement in educational videos for online courses this would refer to students' active participation while watching and interacting with video content. This includes actions like starting and finishing videos, pausing or rewinding to review concepts, taking notes, answering embedded questions and completing related tasks. High engagement often leads to better understanding and retention of material. Instructors can enhance this engagement by incorporating interactive elements, clear visuals, and concise, engaging content in the videos.

### 1.3.2 Emotional Learning Engagement

Emotional learning engagement in educational videos refers to the feelings and attitudes students experience while interacting with the content. Positive emotions, such as interest, enjoyment, or excitement, can enhance motivation and deepen learning, while negative emotions like frustration or boredom can hinder engagement. Emotional engagement can be fostered by using relatable examples, a supportive tone, engaging storytelling, and visually appealing content.

The underlying empirical study of this thesis strongly deals with the emotional impressions of the participants while watching learning videos with a facial analysis software (described in Section 2.6.1) in order to gain an impression of a difference between generated instructors by AI and real instructors in learning videos and therefore to obtain if participants feel emotionally connected to the presenter's material which indicates that they are more likely to stay motivated and persist in completing the course.

In addition, there are a variety of measures in learning environments to increase and promote emotional learning, such as choosing the right learning environment, gamification factors, project-based tasks or campus-wide initiatives concerning course design or pedagogical approaches described in more detail by Struger, Br  nner, and Ebner (2024).

### 1.3.3 Cognitive Learning Engagement

Cognitive learning engagement in learning videos involves the mental effort students invest in processing and understanding the content. It includes activities like focusing attention, analyzing information, applying critical thinking, and connecting new knowledge with prior understanding. Effective cognitive engagement is promoted by videos that encourage active learning (see 1.2.1), such as through challenging questions as well as opportunities for reflection or feedback. When students are cognitively engaged, they are more likely to grasp complex concepts, retain information, and apply their learning in meaningful ways.



### 1.3.4 Measurement of Learning Engagement

Tracking these behaviors can help measure students' involvement and course effectiveness and predict possible learning outcomes. Findings of Struger, Br  nner, and Ebner (2024) show some of these possibilities for measuring learning engagement in higher education:

- **Metadata**  
Based on the *Learning Management System (LMS)*, teachers have the opportunity to access various collections of student data and structure them in a meaningful way to obtain students' learning engagement (e.g. registration times, learning progress with completion rate, time spent, document access, video views, chat tracing etc.).
- **Objective Measurements** (*used for this study*)  
Objective measurement is something that is measured consistently. There are no other factors that can alter the data gathered with this measurement (e.g. how well someone can perform a set number of tasks).
- **Gamification Indicators**  
Gamification serves as a very prominent approach to implement game mechanics in non-gaming environments to enable more interactive learning experiences. Such learning environments enable teachers to obtain useful insights based on different gamification elements used (e.g. certificates, badges, points, levels, etc.) and therefore measure students' learning engagement effectively.
- **Affective Computing** (*used for this study*)  
Affective computing is a technology that uses artificial intelligence to recognize human affects and emotions through computers such as face reader or eye tracker software to obtain students' attention or emotional responses regarding tasks, texts or even videos.
- **Personal Feedback** (*used for this study*)  
Another way to measure students engagement is to provide specific questionnaires, immediate online surveys or intern feedback channels.

### 1.3.5 Digital Inclusion

In the context of learning engagement and imparting enhanced learning, digital inclusion is also playing an increasingly important role. Access to learning content regardless of language, location and social status also has a great impact on the learning engagement for a large groups of students.

In particular, in the future, more and more emphasis will be placed on students with disabilities. Through **Adaptive Learning Environments**, already mentioned in 1.1.2, educational content can be made available in an individually customizable manner (e.g. by subtitling in videos and the conversion of audio to text) where also the behaviour of the *Learning Management System (LMS)* itself can be adapted through interfaces, course flow, organization or difficulty to specific student models. Recent studies such as Murtaza et al. (2022) deals with issues, challenges and possible solutions regarding AI based personalized e-learning systems.

However, according to Ingavélez-Guerra et al. (2022) there is still a lack and frequently detected barriers in accessibility determined in current common online learning environments from the own experience of students with disabilities in virtual environments which is due to lack of implementation of accessibility regulations in educational resources and learning objects increasing the necessity for ideal implementations through automatic tools and further research on this field in the future for contributing to students' overall learning engagement and positive learning experiences.

## 2 Research Method

Given the recent rapid progress of generative AI and increasing demand of multimedia based learning content, this thesis is based on an empirical study investigating the use of AI-generated virtual instructors presenting educational content and revealing effects on learning behaviour and emotional engagement of students' learning experience. According to current state of research on this field, there is one study of Pataranutaporn et al. (2022) comparing different AI-generated characters and their influence on participants motivation towards learning. Another related study by Leiker et al. (2023) investigated the potential of synthetic learning videos comparing learning impacts between virtual and human avatars in learning videos. However this study's main emphasises focus on students' learning behaviour in combination with emotional responses and personal attitude during the comparison of virtual and real instructors in learning videos.

### 2.1 Participants

This study included (N=55) students, teachers and adult learners, recruited from courses at Technical University Graz<sup>1</sup>, a graduating class of The Federal Upper Secondary School Monsberger Graz<sup>2</sup> as well as students and graduates of the University of Klagenfurt (AAU)<sup>3</sup>, University of Applied Science<sup>4</sup> and the Viktor Frankl Teacher Training College in Carinthia<sup>5</sup>. The age range of the participants was from 17 to 62 with an average age of 25.3

---

<sup>1</sup><https://www.tugraz.at/home> (last access November 2024)

<sup>2</sup><https://www.borg1.at/> (last access September 2024)

<sup>3</sup><https://www.aau.at/> (last access September 2024)

<sup>4</sup><https://www.fh-kaernten.at/> (last access September 2024)

<sup>5</sup><https://www.ph-kaernten.ac.at/> (last access September 2024)

## 2 Research Method

years. Of the participants, 55.5% identified as male and 44.5% identified as female. With regards to education, the majority with 38.18% are enrolled as a teaching student in various subjects, 21.82% prospective students, 16.36% graduates or undergraduates in technical subjects, 14.54% employed teachers and 5.5% held a technical doctorate degree. Additionally, regarding the general use of AI, 63.64% stated that they already had prior knowledge in the form of text, audio or video generation while 36.36% identified as being unfamiliar or had little experience with the subject matter.

### 2.2 Video Creation

The main emphasis of this study is the introduction of AI-generated virtual avatars in learning videos used in learning management systems based on MOOCs. For this purpose instructor videos, produced using traditional recording methods and capture equipment (see Figure 2.1), were created.



Figure 2.1: Video creation using a Panasonic GH5s camera with Olympus 12-200mm lens with 4k, 25fps, ISO 400, 1/50 shutter speed and 5.6 aperture considering 2m distance to the well-lit green screen and 2m to the front camera.

## 2 Research Method

The learning content of the videos were provided via text-based scripts regarding the topic of educational online platforms. To ensure unbiased results concerning social cues like gender and voice, the videos were created with one female and one male instructor respectively. Both versions were used as a control for the experiment and as the source material for generating the synthetic videos with the virtual avatars. For the creation of the synthetic versions of the instructors and their voice, HeyGen<sup>6</sup> was used to generate text-to-videos (TTV) with the provided text scripts illustrated in Figure 2.2.

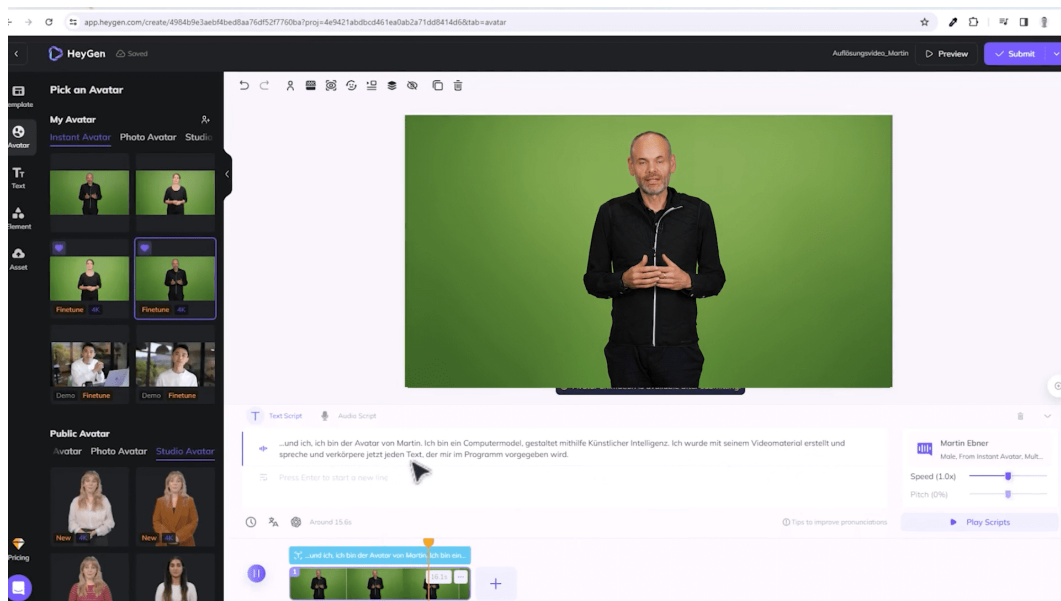


Figure 2.2: Video creation using the AI-powered video tool HeyGen provided with footage of the real instructor.

To ensure photo realistic quality and meaningful comparison between virtual and real avatars in learning videos, the online AI-powered video tool HeyGen was used and provided with footage of the real instructors to establish synthetic clones. Using neural video synthesis, gestures and movements were added to the generated video representation of the instructors driven by TTV input and voice cloning for the final synthetic video result. This process of video creation by HeyGen can be applied to any custom footage

<sup>6</sup><https://www.heygen.com/> (last access November 2024)

## 2 Research Method

or using predefined AI avatars for generating synthetic instructor videos. Two AI-generated avatars, representing one female and one male instructor respectively, each presenting the two provided educational content descriptions were created in the course of this study resulting in four different synthetic videos serving as the experiment sequences compared to control sequences with real footage of the instructors illustrated in Figure 2.3.



Figure 2.3: From left to right: Video frame with the real and AI-generated presenter of the male instructor used for the study, created in Adobe Premiere Pro.

### 2.3 Experiment Setup

Using one video with a synthetic representation of the instructor and one video with the real instructor representation, two micro-learning courses with video lengths of 2-5 minutes were designed for the experiment. The main goal of the courses was to provide learners with an easy-to-complete learning unit explaining the basics of two different online learning platforms. Every course consisted of the corresponding learning video followed by two single choice questions regarding the content and mentioned facts.



## 2 Research Method

Each participant was presented with two courses where the first course always served as the experiment sequence including the virtual avatar while the second course was presented by the real instructor from the other gender respectively as the control sequence. Afterwards a final video for clarification and revealing, that the first video was generated using artificial intelligence, with a video length of about 45 seconds was presented to the participants promising strong emotional response and personal reflection during watching.

To avoid biased results regarding social cues like gender or voice during watching the learning videos, topics and instructors were switched after each experiment run. After the video session, participants were interviewed for about 10 minutes to capture personal impressions, opinions, feelings and grounds for suspicion during watching the videos in a structured interview. The whole experiment was conducted in presence in soundproof quiet rooms considering stable lighting conditions illustrated in Figure 2.4.



Figure 2.4: Experiment setting at the educational technologies organizational unit of TU Graz.

## 2.4 Implementation

The implementation of the research method illustrated in Figure 2.5 shows the process and which steps in which order were necessary to answer defined research questions of this thesis.

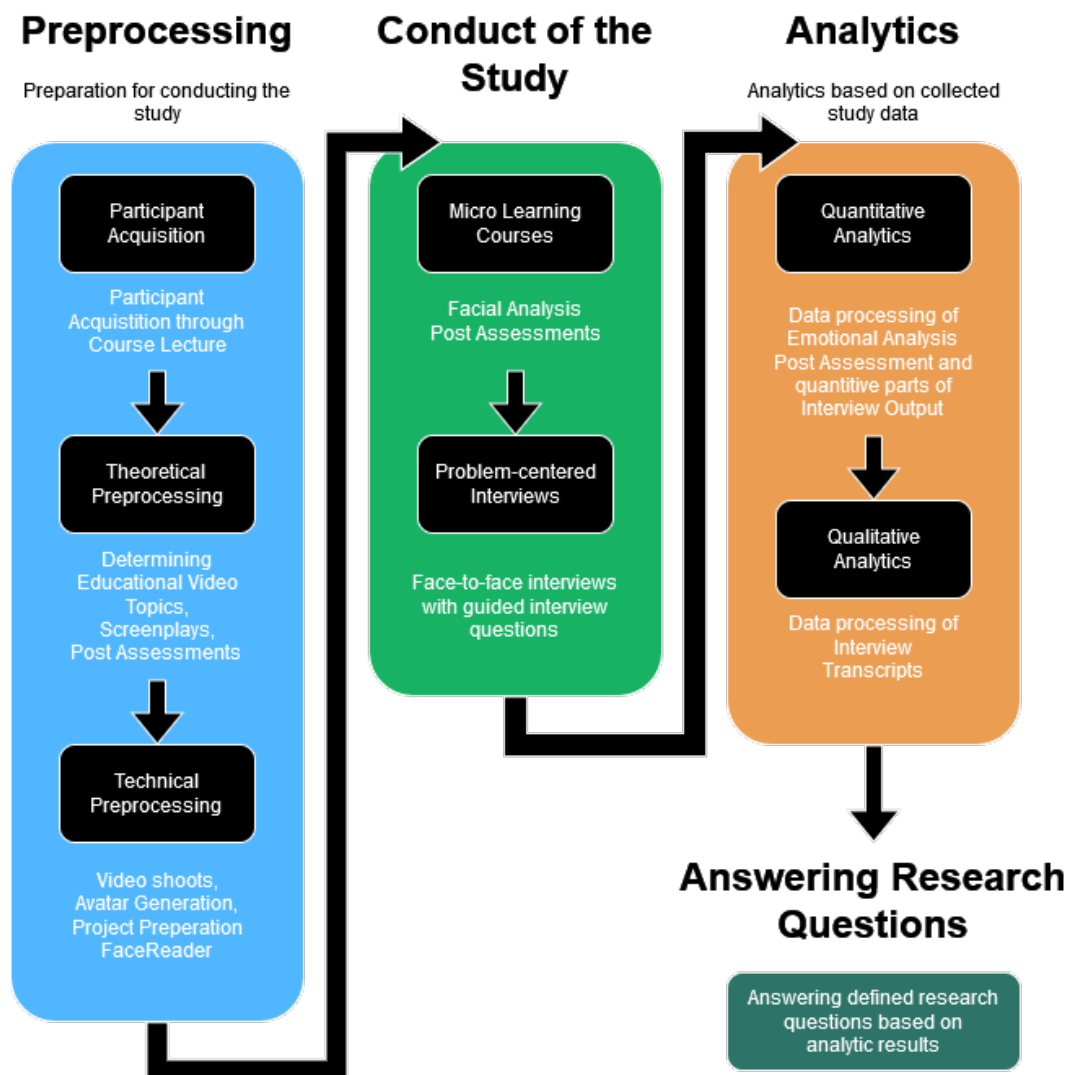


Figure 2.5: Flow chart of the process of study implementation.



## 2.5 Course Design

Figure 2.6 illustrates the course design of the presented learning courses created and used for this study. Each participant was presented with one micro course in an evenly distributed ratio over all participants to reduce biases regarding gender of the presenter and course topic. After each learning video participants were asked to answer two single-choice questions regarding the past course topic. In the end the stimulus sequence was presented revealing that the first video was generated using AI. All individual videos are published under a CC BY license in the repository of TU Graz at <https://doi.org/10.3217/ggn-cr-sg773>.

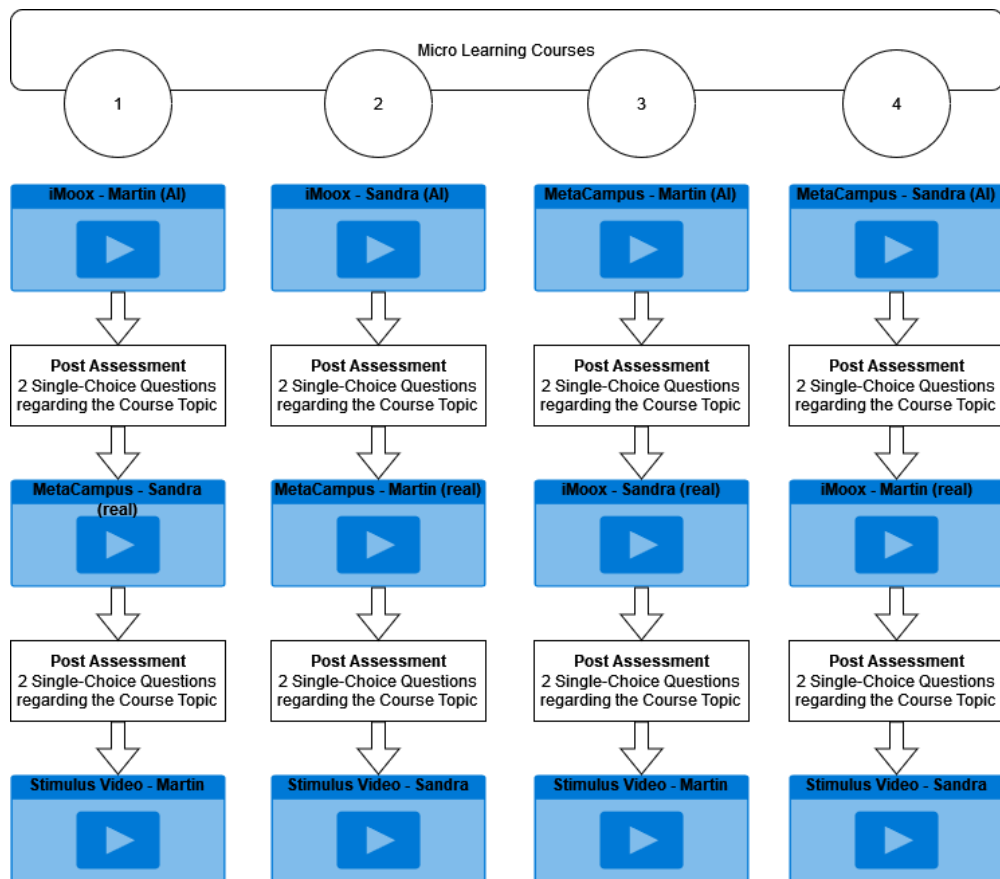


Figure 2.6: Structure of micro learning courses designed for this study.

## 2.6 Evaluation

### 2.6.1 Facial Analysis

The first part of the evaluation concentrated on the attention, responsiveness and emotional response of the participants regarding learning engagement and learning outcome. During each experiment run, participants were recorded through a webcam and analyzed regarding their corresponding facial expressions with the facial analysis tool FaceReader Online<sup>7</sup>. The online tool is capable of capturing emotions by analyzing facial expressions and runs on a cloud server which was the more convenient option for the experiment setup instead of a local installation. Each course video was analyzed with respect to 7 basic emotional expressions and metrics like valence, arousal, responsiveness and attention separately. Analysis and chart generation was carried out with the online platform and python scripting.

### 2.6.2 Post Assessment

For evaluation purposes concerning instructor influence and students' learning outcome, answers to the corresponding course questions were analyzed through spreadsheets and python scripting for visualization purposes.

### 2.6.3 Problem-centered Interview

The second part of the experiment evaluates personal impressions, triggers, feelings and opinions of using AI based technologies in learning environments in general. For this purpose working with problem-centered interviews through voice recording in combination with qualitative content analysis described by Mayring and Fenzl (2019) served as the most promising methodology. First drafts of the voice recording transcriptions were created with the automatic speech recognition (ASR) system Whisper<sup>8</sup>.

---

<sup>7</sup><https://facereader-online.com> (last access November 2024)

<sup>8</sup><https://openai.com/index/whisper/> (last access November 2024)

## 2 Research Method

For this purpose, a problem centered interview consisting of 3 topic blocks with 1 key question and 3-4 sub-questions each was prepared and analyzed following the inductive category formation steps illustrated in Figure 2.7.

### Topic Block 1 - Perception of AI Instructor

- **Would you have thought that the first video was created with an AI avatar?**
  - Did the AI avatar seem strange or unnatural to you?
  - Do you know the person in the video personally?
  - Was there a particularly striking moment regarding the learning videos?
  - Have you heard of this technology?

### Topic Block 2 - Learning videos with AI software

- **Can you learn equally well with the AI avatar?**
  - Did you notice a change in your perception or concentration between the two videos?
  - Were you able to remember information or understand content in the video just as well as the AI avatar?
  - Would you rather see a real person in a learning video?

### Topic Block 3 - Future of learning videos

- **Would you like to see this technology used in the future?**  
**Why / Why not?**
  - If you are watching an educational video in another language, would you prefer a video with text translation in your language or a video translated using AI?
  - Which avatar would you rather see in a learning video? Yourself, the teacher, a friend or a cartoon character?
  - What advantages and disadvantages do you personally see in using AI in learning environments?
  - Would you like to have an AI avatar created of yourself?

## 2 Research Method

The original version of the interview questions is also provided in [A.2](#).

As part of the anonymized evaluation, additional metadata of the participants, such as age as well as the field of study and level of education to assess the technical background knowledge, were recorded (on a voluntary basis) at the end of the interview.

### Topic Block 4 - Metadata

- How old are you?
- What is your highest level of education
- What field of study are you currently or have you been studying?

All study participants were informed about the anonymized data analysis and subsequent evaluation. Furthermore, each individual participant signed a consent form for the recording and evaluation of auditory (interview recordings) and biometric data (facial recognition) as well as their legally compliant storage, processing and deletion before participating in the study accordingly as provided in [A.1](#).

## 2 Research Method

The problem-centered interviews in the last step of the survey were analyzed following the process model of the inductive category formation described by Mayring and Fenzl (2019) illustrated in Figure 2.7 below. Step 5-8 describe the process of category formation in an iterative manner.

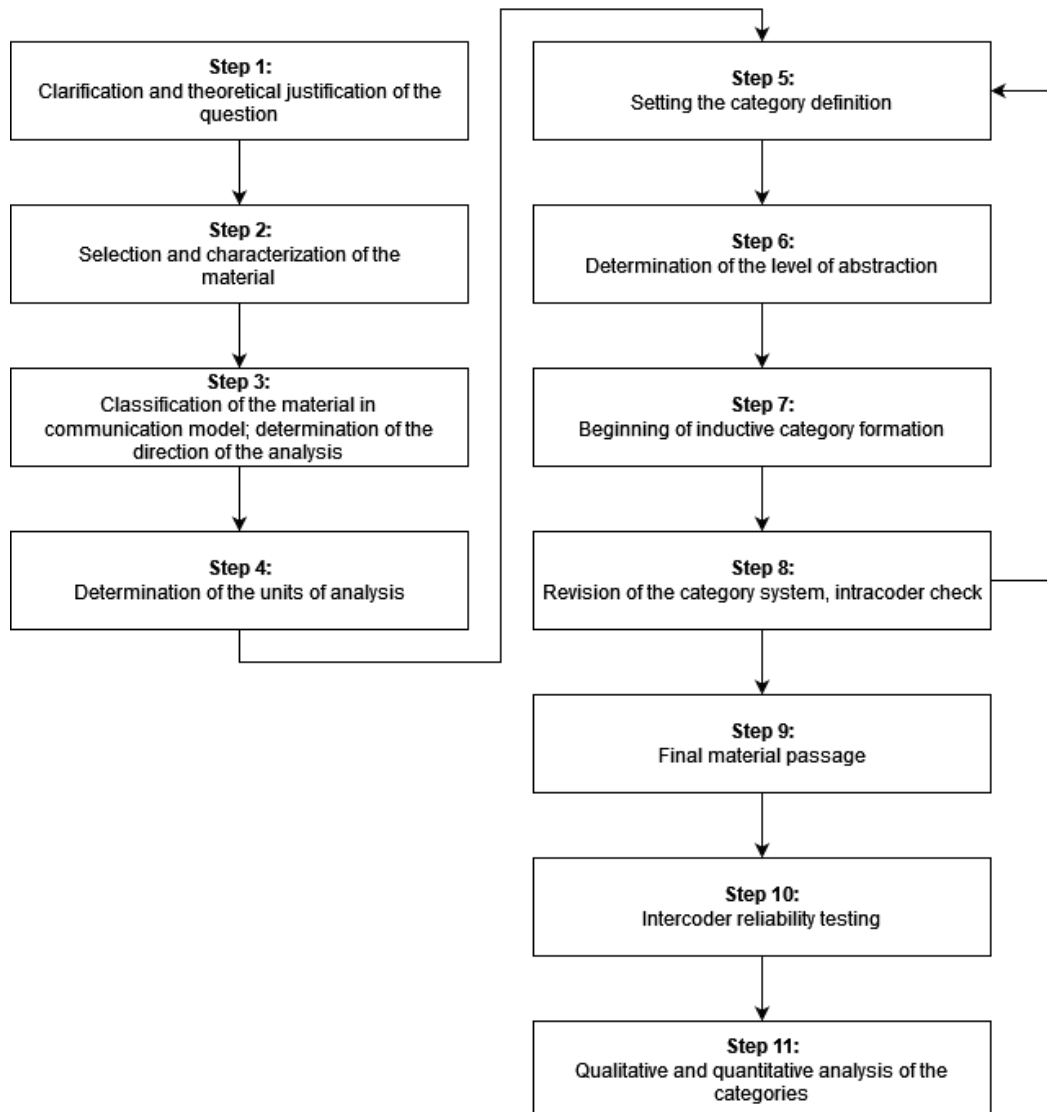


Figure 2.7: Process model of inductive category formation adapted from Mayring and Fenzl, 2019.

## 2 Research Method

The qualitative analysis of the face-to-face interviews was carried out following the inductive category information described by Mayring and Fenzl (2019) which is based on a summary approach developed through an intrinsic methodology of the used material. This methodology uses the given text-based material as the core input source for the category formation where coding guidelines are extracted from the material step by step in an iterative manner.

However, for the sake of completeness, the category formation can also be done with a structural approach calling the deductive category formation. This method follows an extrinsic approach based on specific predefined theoretical frameworks, meaning that conditions for the category formation are extracted by existing coding guidelines from scientific papers with similar context of the predefined research question.

The main goal and output of the qualitative content analysis (in both approaches) is the reduction of the whole content to the essentials. The aim is to create a category system that abstractly represents the material and to relate these categories to one another in the form of higher-level and lower-level categories.

To implement the process model of the inductive category formation illustrated in Figure 2.7 on a practical approach, each step has to be defined for the content analysis properly:

- **Step 1:** For clarification and theoretical justification of the question, which represents the basis of the analysis, the underlying research question for the interviews of this thesis was defined as follows:

**"Can educational content be learned and understood just as well with the help of generated AI lecturers in learning videos as with real lecturers in learning videos?"**

## 2 Research Method

- **Step 2:** The selection and characterization of the material, the interview questions were structured beforehand which means that all information regarding a specific topic in the interview was designed in a guided manner (see 2.6.3):

```
<Topic Block>  
<Central Guiding Question>  
<Sub-questions>
```

- **Step 3:** For classification of the material in communication model, the representing transcript of each participant of the face-to-face interview was used. The material had to be uniformly structured as follows:

```
<Interviewer (indicated by the abbreviation "I")>:  
<Question (ends with "?")>  
<Participant (indicated by the abbreviation "P")>:  
<Answer>
```

- **Step 4:** For the determination of the units to be analyzed, the whole material, including all transcripts with complete statements of each participant (n=41) was used.
- **(Iterative) Step 5:** For setting the category definition, the first subject areas were derived and predetermined by the structure of the interview itself. Three areas (*Perception, Learning Experience and Future of AI in Learning Environments*) emerged from the interview design for first orientation of the category formation.
- **(Iterative) Step 6:** For the determination of the level of abstraction each complete statement was reduced to the essentials and separated into keywords as follows:

```
Abstract Level 0: "Since there were no differences  
between the videos and I didn't notice it, I would  
say yes. It's absolutely equally easy."  
Abstract Level 1: "no differences between the videos"  
Abstract Level 2: "no differences"
```

## 2 Research Method

- **(Iterative) Step 7:** Beginning with the category formation starting with predetermined categories based on the interview structure, by forming additional top-level and sub-level categories or by merging of already existing top-level or sub-level categories.
- **(Iterative) Step 8:** The revision of the category system is recommended for a throughput of 10-30% of the total material. In this case a revision was carried out at around 30% of the processed material. At this point, category formation and definition had to be checked to cover the predefined research questions and abstractly representing the whole material properly.
- **Step 9:** Final material passage of the whole material to be analyzed. In this case all transcripts were analyzed in whole with final adjustments concerning the process of the inductive category formation.
- **Step 10:** At this point a reliability test can be done with a second independent run by another person or tool like QCMap<sup>9</sup> or even ChatGPT<sup>10</sup>. Considering the topic of this scientific work this step was carried out with ChatGPT with regards to category formation only following recommended prompting guidelines described by Zhang et al. (2024).
- **Step 11:** The qualitative and quantitative analysis of the content was completed by the frequency analysis of individual statements represented by the formed categories. This analysis underlines the strength of statements and categories throughout the whole material and helps to derive meaningful conclusions from the analysis.

---

<sup>9</sup><https://www.qcamap.org/ui/de/home> (last access November 2024)

<sup>10</sup><https://chatgpt.com/> (last access November 2024)



## 3 Results

### 3.1 Post Assessments

To properly answer Research Question 1, participants' responses to post-learning assessments were scored and compared with experimental and control sequences to test if the use of AI avatars affected learners knowledge acquisition. The two-way frequency table 3.1 describes the descriptive results with respect to score and portion of wrong and right answers of the knowledge test after each learning course differentiated by joint and marginal frequencies. Joint frequencies are the counts that show how often combinations of two variables occur together while marginal frequencies represent the overall frequency for one variable, regardless of the other variable.

		Answers — Virtual Instructor							
		Right		Wrong		Not Finished		Total	
		n	%	n	%	n	%	n	%
Answers — Real Instructor	Right	72	65.45%	10	9.09%	2	1.82%	84	76.36%
	Wrong	19	17.27%	3	2.73%	0	0%	22	20%
	Not Finished	4	3.64%	0	0%	0	0%	4	3.64%
	Total	95	86.36%	13	11.82%	2	1.82%	110	100%

Table 3.1: Frequency and percentage of answers to post assessments regarding courses with virtual and real instructors

Considering a total number of 2 questions per micro-course with (N=55) participants, 110 questions per course were answered in total. 2 answers to questions regarding videos with virtual presenter and 4 answers of the real presenter could not be obtained by the online tool and gave no insights. Therefore the experiment ended up with a total of 108 answered

### 3 Results

questions for the avatar course and 106 for the real course. With controlled alternating topics, meaning both datasets for the courses ending up with the same occurrence of course topics, courses with virtual instructors obtained slightly more right answers and less wrong answers than courses with real instructors.

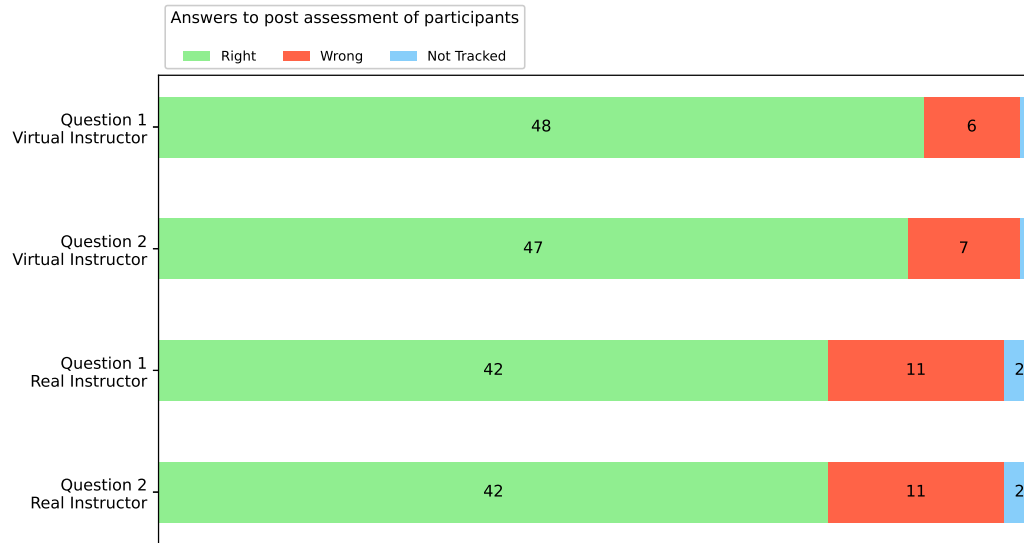


Figure 3.1: Discrete distribution of wrong and right answers to post assessment of learning videos with real instructor and virtual instructor.

The observed results show that students trained by virtual instructors perform 10% better or about 8% less worse, solely regarding the quantitative perspective of the post assessments. The comparison ended up with a total of 95 right (86.36%) answers, 13 wrong answers (11.82%) and 2 untracked records (1.82%) after taking courses showing the virtual instructor in contrast to 84 right (76.36%) and 22 wrong answers (20%), excluding 4 untracked records (3.64%) by the face reader software, after watching courses with real instructors. Questions of the post assessments for the respective course topics are provided in [A.3](#).

### 3 Results

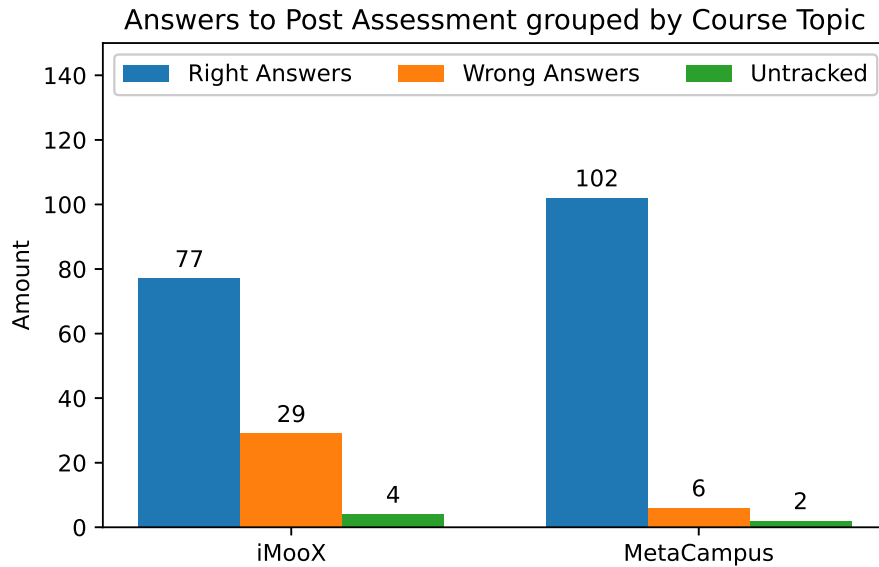


Figure 3.2: Participants' answers to post assessments grouped by the two course topics "iMooX" and "MetaCampus".

Regardless of the type of instructor in the learning videos, another interesting observation during the analysis process of the post assessment was the distribution of right and wrong answering of questions from participants regarding the two defined course topics, illustrated in Figure 3.2. During the interviews (see section 3.3), some participants stated that they were able to concentrate better on the content because they already had some prior knowledge of the topic. However, these participants stated that they had already dealt with iMooX<sup>1</sup> during their studies or work, whereas MetaCampus<sup>2</sup> was a relatively unknown topic for most participants. However, this statistic is not reflected in the numbers illustrated in Figure 3.1, since both course topics were equally distributed between real and artificially generated lecturers in the learning videos.

<sup>1</sup><https://imoox.at/mooc/> (last access November 2024)

<sup>2</sup><https://metacampus.unite-university.eu/> (last access November 2024)

### 3 Results

## 3.2 Emotional Classification

To establish a connection between participants' answers to post assessment questions and their emotional inventory during the learning videos, results from FaceReader Online were extracted covering each experiment with the corresponding experimental and control sequence using comparable group results. For this purpose the share of emotional states during both, videos with real (see Table 3.2) and virtual instructor (see Table 3.3), were measured and compared to each other illustrated in Figure 3.3. The emotional states were broken down into "Neutral", "Happy", "Sad", "Angry", "Surprised" and "Disgusted" turning out as the most common ones. Possible classifications like "Scared" and "Contempt" were excluded from comparison due to slight contribution (threshold of 1%) to the overall outcome.

Real Instructor	Percentage of Emotional States					
	Neutral	Happy	Sad	Angry	Surprised	Disgusted
MetaCampus Sandra	0.728	0.020	0.073	0.054	0.006	0.005
MetaCampus Martin	0.780	0.019	0.018	0.020	0.010	0.032
iMooX Sandra	0.890	0.003	0.024	0.033	0.021	0.001
iMooX Martin	0.846	0.004	0.023	0.021	0.023	0.001
	0.811	0.0115	0.0345	0.032	0.015	0.00975

Table 3.2: Share of emotional states of learning courses with real instructor.

Virtual Instructor	Percentage of Emotional States					
	Neutral	Happy	Sad	Angry	Surprised	Disgusted
MetaCampus Sandra	0.875	0.011	0.021	0.014	0.012	0.001
MetaCampus Martin	0.921	0.005	0.014	0.026	0.020	0.002
iMooX Sandra	0.775	0.075	0.024	0.018	0.007	0.037
iMooX Martin	0.797	0.005	0.079	0.041	0.005	0.007
	0.842	0.024	0.035	0.025	0.011	0.012

Table 3.3: Share of emotional states of learning courses with virtual instructor.

### 3 Results

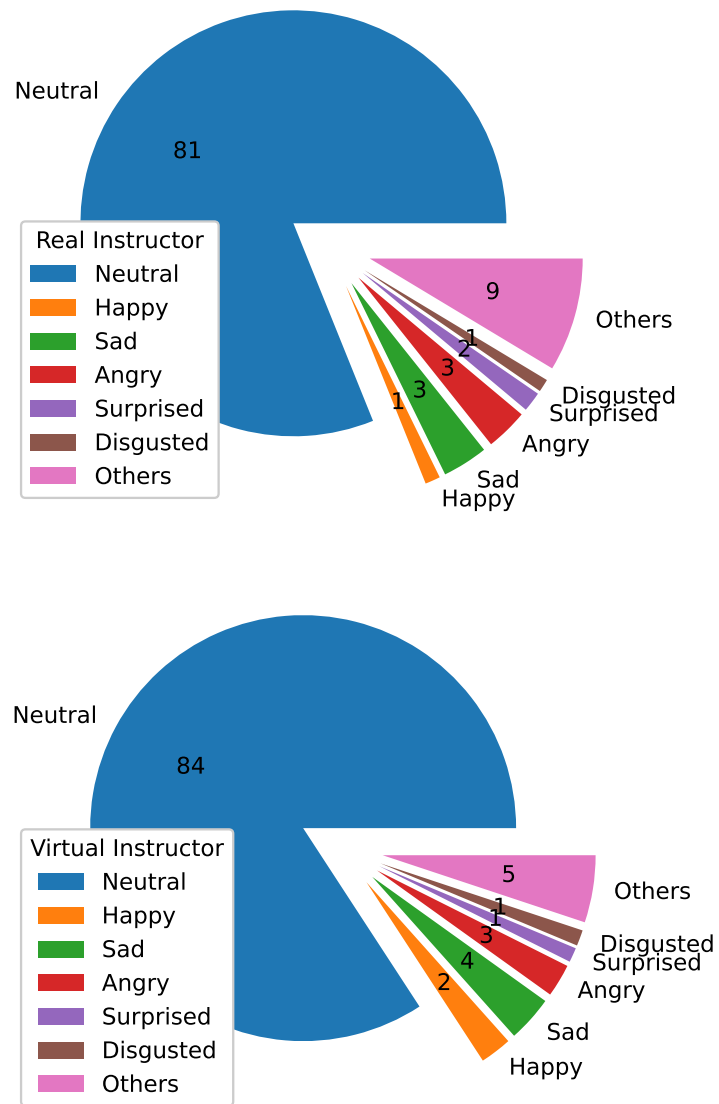


Figure 3.3: Comparison of captured emotional states using FaceReader software of learning videos with real instructors (top) and virtual instructors (bottom) in percent.

### 3 Results

The used facial analysis software FaceReader Online<sup>3</sup> also internally evaluates the quality of the recorded participants during the classification of emotions on a scale of 1 (not good) to 10 (very good). Care was taken to ensure that the setting for the recordings can be carried out with constant conditions across all participants (see Figure 2.4). Figure 3.4 illustrates that over 85% of the recordings (N=55) were rated with a quality factor of 8, one recording with 9, a few with 7, 1 recording with 6 and 1 recording with the worst quality ranking with 4.

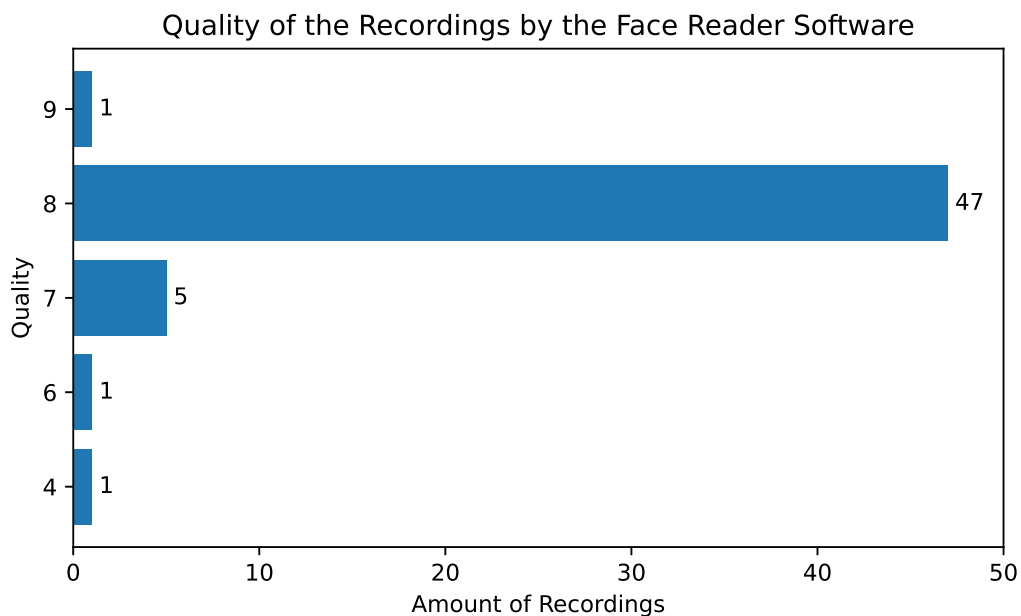


Figure 3.4: Quality ranking of the facial analysis software regarding recordings of the participants.

Upon closer inspection of the varying quality of the recordings, it was noticeable that recordings of poorer quality came from participants who wore glasses, were too close to the screen or were squinting.

---

<sup>3</sup><https://facereader-online.com/> (last access November 2024)

### 3 Results

In addition to get a deeper insight into both learning courses and emotional classification during watching, control and experimental sequences were compared to each other including the most noticeable metrics during the analysis illustrated in Figure 3.5 and 3.6 for the first course topic "iMooX" and Figure 3.6 and 3.8 for the second course topic "MetaCampus" with distinction between male and female instructors respectively.

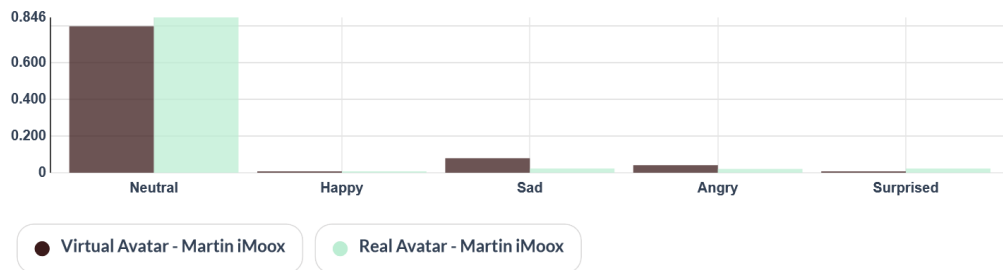


Figure 3.5: Comparison of classified emotions between real and virtual male instructor of the first learning course.

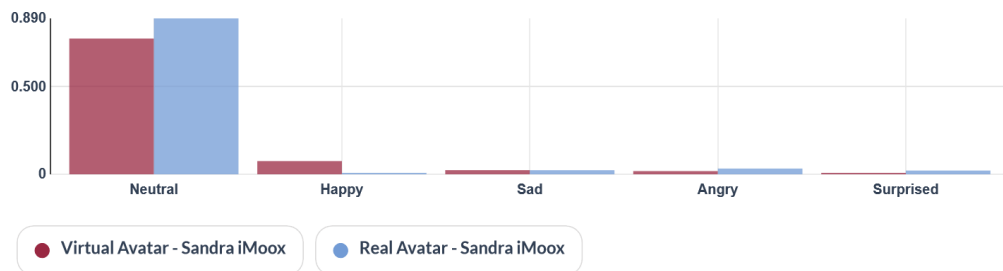


Figure 3.6: Comparison of classified emotions between real and virtual female instructor of the first learning course.

The classification of emotions during the first learning course reveals stronger neutral feelings in case of the real instructor video regardless the gender role. Feelings like happiness, sadness or fury show a higher peak regarding the first learning course in case of the virtual instructors in general.

### 3 Results

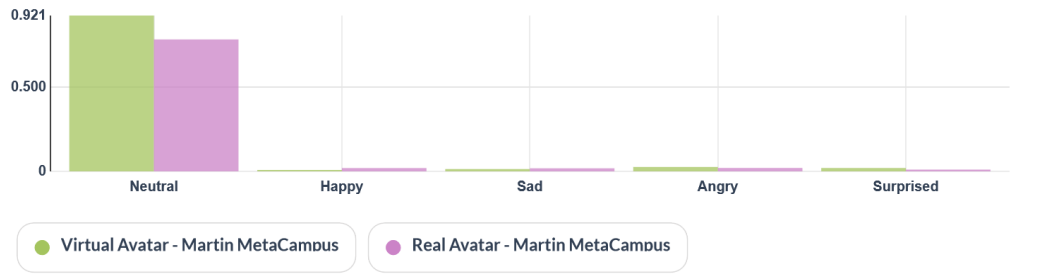


Figure 3.7: Comparison of classified emotions between real and virtual male instructor of the second learning course.

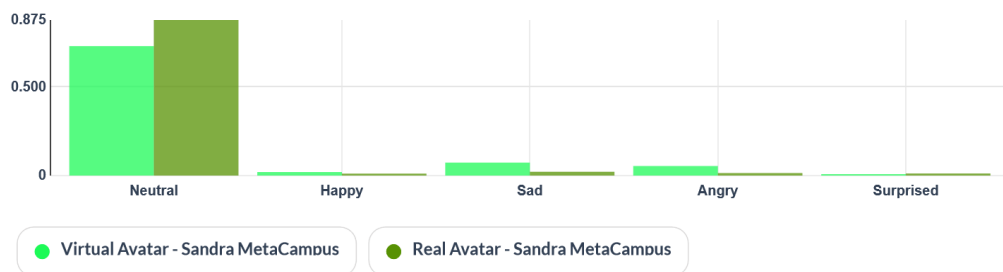


Figure 3.8: Comparison of classified emotions between real and virtual female instructor of the second learning course.

Classification of emotions during the second learning course reveals opposite values of neutral feelings comparing female and male instructors and their corresponding avatars. Happiness, sadness and fury show a slightly higher peak in case of the real male instructor (see Figure 3.7) and less in case of the female instructor (see Figure 3.8) compared to the corresponding virtual instructor.



### 3 Results

In addition to the classification of emotional states, measurements regarding students' responsiveness, valence, arousal and attention score were used for the comparison. The responsiveness<sup>4</sup> score describes how sharp changes of the emotional states of participants occurred in regards with the corresponding educational content. The two basic related emotional dimensions valence and arousal correspond to the emotion polarity (positive, negative or neutral) of each learning course as well as the power of emotions which are described by Elamir, Al-Atabany, and Eldosoky (2019) in more detail and illustrated in Figure 3.10. In order to learn, the act of directing the mind to listen, see, or understand was tracked via the attention score during watching the learning videos.

Real Instructor	Valence		Arousal		Responsiveness		Attention	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
MetaCampus Sandra	-0.094	0.172	0.017	0.009	0.166	0.169	1.000	0.002
MetaCampus Martin	-0.037	0.111	0.021	0.016	0.105	0.102	1.000	0.000
iMooX Sandra	-0.027	0.072	0.025	0.020	0.084	0.085	0.999	0.002
iMooX Martin	-0.015	0.059	0.018	0.012	0.088	0.074	0.998	0.005
	-0.04325	0.1035	0.02025	0.01425	0.11	0.1075	0.99925	0.00225

Table 3.4: Mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of valence, arousal, responsiveness and attention of learning courses with real instructor.

Virtual Instructor	Valence		Arousal		Responsiveness		Attention	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
MetaCampus Sandra	-0.011	0.037	0.015	0.013	0.074	0.073	1.000	0.000
MetaCampus Martin	-0.012	0.047	0.014	0.008	0.070	0.080	1.000	0.000
iMooX Sandra	0.014	0.222	0.029	0.033	0.168	0.210	1.000	0.001
iMooX Martin	-0.107	0.134	0.020	0.016	0.142	0.144	1.000	0.001
	-0.029	0.11	0.0195	0.0175	0.1135	0.12675	1.000	0.0005

Table 3.5: Mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of valence, arousal, responsiveness and attention of learning courses with virtual instructor.

<sup>4</sup>[https://en.wikipedia.org/wiki/Emotional\\_responsivity](https://en.wikipedia.org/wiki/Emotional_responsivity) (last access September 2024)

### 3 Results

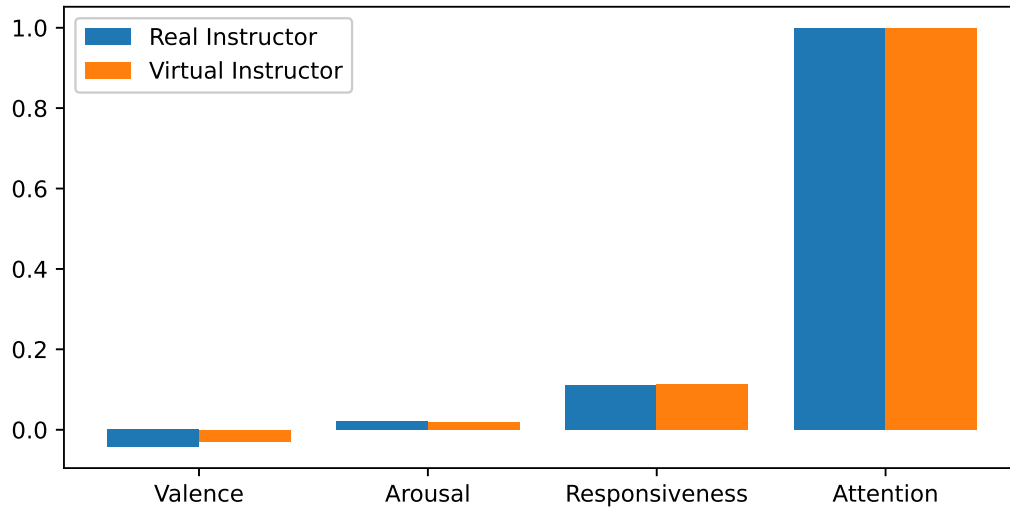


Figure 3.9: Differences between valence, arousal, responsiveness and attention during the micro-courses comparing virtual and real instructor.

Observing the results illustrated in Figure 3.9 reveals slightly more negative valence and higher arousal in videos with the real instructors while participants' responsiveness during watching the videos indicate no differences between virtual and real instructors. The attention score refers to the eye contact with the screen which is almost 100% since participants were advised to pay attention to the screen permanently because of the recording quality.

### 3 Results

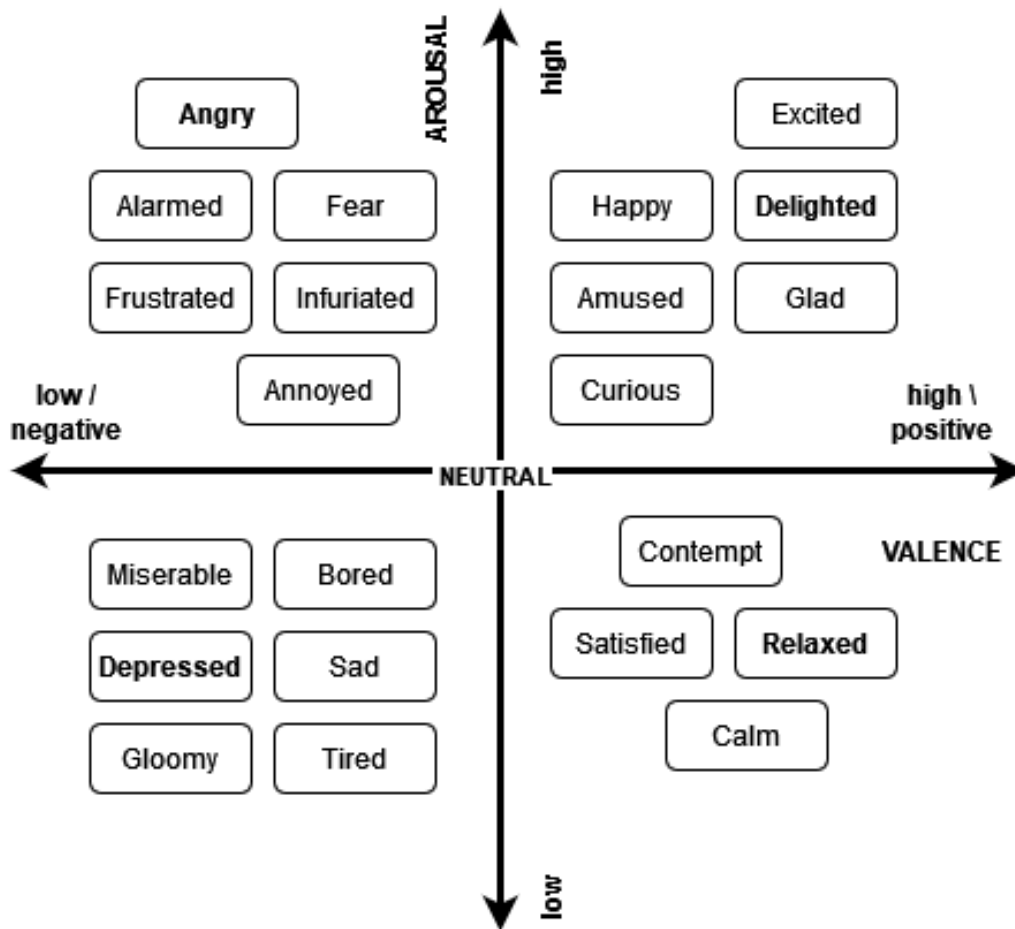


Figure 3.10: The Arousal-Valence Model of Emotions adapted from Elamir, Al-Atabany, and Eldosoky (2019).

The arousal-valence model is a psychological framework used to describe and categorize emotions. The model consists of two dimensions:

- **Arousal:** Refers to the intensity or energy level of an emotion, ranging from low (calm, relaxed) to high (excited, tense).
- **Valence:** Refers to how positive or negative the emotion feels, ranging from unpleasant (sadness, anger) to pleasant (happiness, joy).

### 3 Results

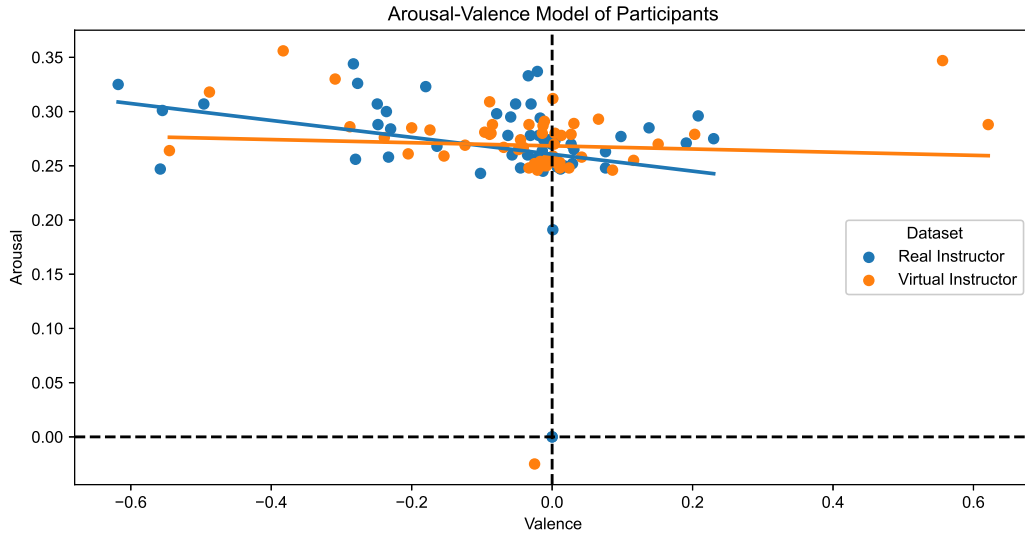


Figure 3.11: The Arousal-Valence Model of (N=55) participants.

For the model (N=55) participants were considered while (n=2) participants returned no result sets from the face reader software for the valence and arousal values in the real instructor dataset (marked blue). For the sake of analysis, average values of the entire corresponding learning video with real or virtual instructors were taken into consideration.

Compared to the given reference model 3.10, participants' emotions indicate a neutral state in general. Some outliers can be observed with positive arousal and negative valence values indicating annoying or angry moods more likely while other value pairs show both positive arousal and valence indicating more excited or happy emotions. The distribution between the dataset of real instructors and the dataset of virtual instructors resulted in slightly different value pairs and corresponding regression lines.

An example of an individual result view with all metrics concerning the emotional classification through FaceReader Online is provided in A.4.

### 3.3 Problem Centered Interview

Following the analysis of the micro-courses and their quantitative aspects concerning post assessment and emotional classification, the survey participation concluded with a face-to-face interview mentioned in 2.6.3 regarding the usage of AI techniques in learning environments. The analysis of the interview serves as both, qualitative and quantitative component of the survey and should give useful insights about the perception and usage of AI in learning environments, qualitative improvement, general acceptance and individual learning experience in more detail.

For the qualitative part of the interview analysis, an adapted process model of inductive category formation by Mayring and Fenzl (2019) was used as the underlying analysis framework, illustrated in Figure 2.7 and described step by step in 2.6.3. In the course of the evaluation, following categories were inductively developed for abstractly representing the content of the interview transcripts (n=41):

- Perception
- Learning Experience
- Advantages of AI in Learning Environments
- Disadvantages of AI in Learning Environments
- Threats about AI
- Personal Perspective
- Future of AI in Learning Environments

### 3 Results

Perception			
Category	Description	Example	Frequency
Conspicuousity of the AI avatar	This category describes conspicuous details of the AI avatar in the learning video, in terms of gestures, facial expressions and tone modularity.	"I felt it was a bit too repetitive." "The pauses in speech were often in the wrong place. [...] The modulation of the voice was just not right."	3 <sup>1</sup>
Perception of the AI avatar	This category describes how the participant perceives the AI avatar in the learning video, especially in terms of naturalness and progress in technology.	"I didn't realize at first that this wasn't a real person. That's why I was able to understand things right away."	11
Artificiality of the AI avatar	This category describes how artificial the AI avatar seemed for some of the participants.	"You notice that the avatar is not a real person. The way he speaks and moves feels somehow fake."	6

Table 3.6: Inductive category table how participants perceived the AI avatar with corresponding sub-categories.

Although about 75% of the participants (n=31) conducting the interview after the micro courses stated that they had perceived the AI generated human as unnatural in terms of emphasis, gestures, tone modularity or facial expressions (see Table 3.6), only about 21.8% of all participants (N=55) clearly recognized the AI avatar, 20% had partially suspected it and 58.2% noticed no difference at all from the real lecturer in the learning videos as illustrated in Figure 3.12.

### 3 Results

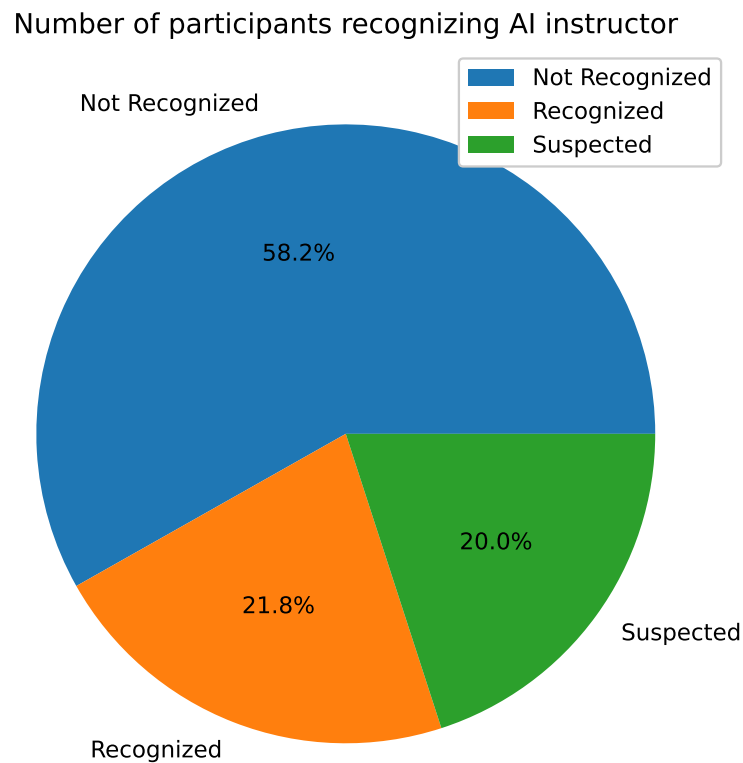


Figure 3.12: Distribution of the amount of participants recognized AI generated instructors.

Since there may also be some deviation regarding the personal knowledge of the person in relation to the perception of the AI avatar, Figure 3.13 illustrates the distribution of participants based on the prior knowledge of the teaching person on identifying the AI generated instructor during the learning videos.

### 3 Results

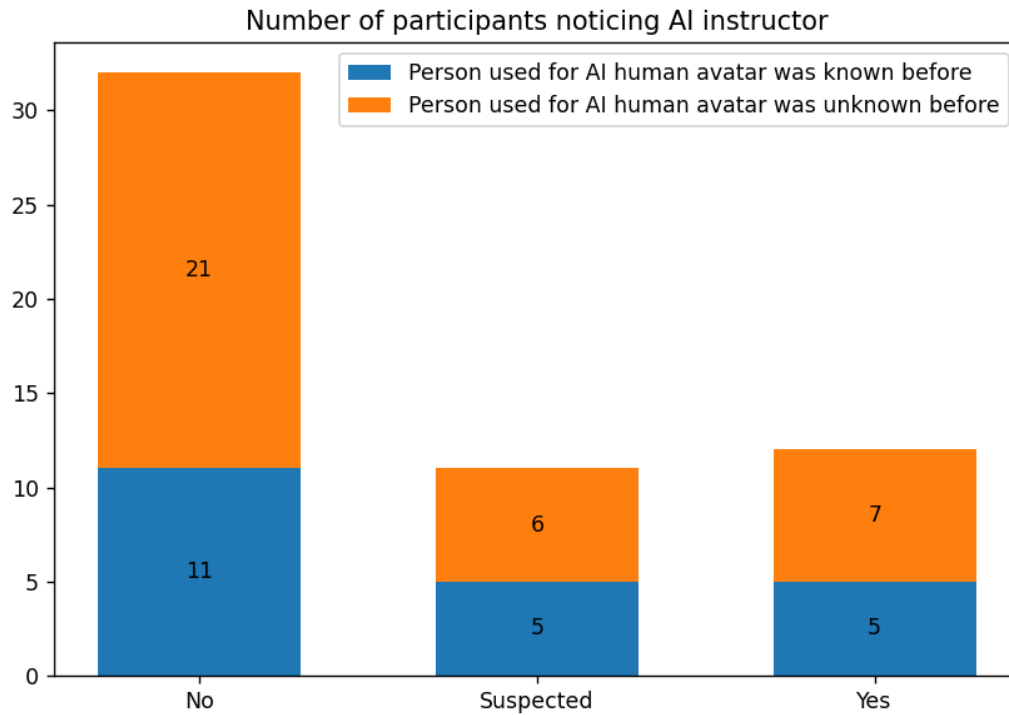


Figure 3.13: Number of participants noticing AI instructors grouped by prior knowledge of the corresponding real instructor.

About 38% ( $n=21$ ) of all participants ( $N=55$ ) already knew the AI generated instructor from online courses (MOOCs), personal interaction or courses at the university. Nevertheless, more than half of these participants could not distinguish the artificially generated speaker from the already known real person. It was also interesting to note that even more people who had never seen the real person before clearly recognized the AI avatar or had already suspected the use of this technology.



### 3 Results

Learning Experience			
Category	Description	Example	Frequency
Linguistically Understanding	This category summarizes how well participants were able to understand the learning content linguistically when an AI avatar was used.	"I was able to understand the content without any problems."	41
Equality of learning	This category asks whether participants found learning with an AI avatar to be as effective as learning with a real person.	"Since there were no differences in this video and I didn't notice it, I would say yes. It's absolutely equally easy."	28
Emotional Learning	This category describes that some participants found the AI avatar less authentic, which affected the learning atmosphere and the bond with the teacher.	"It's still something different and comes across differently. So there's a certain distance that would be different with a real person."	7
Distraction by the AI Avatar	Some participants found the AI avatar distracting because they focused on potential errors or the artificial nature of the avatar rather than focusing on the learning content.	"I realized that I was constantly checking to see if something was wrong, and as a result I wasn't really paying attention to the content."	6
Adaptation to Learning Styles	This category describes that participants have different preferences for learning methods depending on their learning style.	"I'm perhaps more of a reader type, I need more time, slower. [...] When I was watching, I was somehow focused on other things."	5
General Necessity of instructors in learning videos	This category addresses the question of whether participants generally prefer a lecturer in learning videos.	"I generally don't need to see the person speaking in a learning video because it tends to distract me."	4

Table 3.7: Inductive category table how the overall learning experience with AI avatars was assessed with corresponding sub-categories.

Table 3.7 illustrates that most of the participants agreed that they could learn just as well with the AI avatar as with a real lecturer in a video. Also all of the participants could linguistically understand the educational content presented by the AI avatar although many of them commented on abnormalities in tone modularity and unusual pauses in speech as illustrated

### 3 Results

in Table 3.6. The importance of emotional learning is also consciously addressed by some participants, as something seems to be missing in contrast to a real lecturer in a learning video.

Since about 40% of all participants (N=55) recognized or suspected one avatar (see Figure 3.12) to be generated by AI, some of them stated that they were distracted from the learning content by looking for mistakes because something seemed to be not quite right. Other participants on the other hand were generally distracted by the presence of a lecturer in the learning video. A few participants stated that the necessity of the presence of a lecturer in a learning video depends on the topic. For example, for how-to videos or videos with programming content, the entire screen is needed to absorb information and only an accompanying voice of the lecturer is needed.

Another interesting aspect was the consideration of different learning types. Some of the participants stated, that they prefer learning educational content by reading rather than watching a learning video explicitly. For this group of participants, the duration of concentration, visuals and the speed of the knowledge being transmitted play a crucial role. When reading at their own pace, this type of group can achieve better individual learning progress and may even repeat sentences. When watching learning videos, they said that they generally do not rewind the video even if they knew it is possible.

In addition, it was also mentioned that prior knowledge of learning content contributes to the overall learning experience. Those affected stated, for example, that in the micro course for iMooX<sup>5</sup>, there was better absorption due to prior knowledge, in contrast to the micro course on the topic of MetaCampus<sup>6</sup>, which turned out to be new input for most participants.

The length of the learning video plays a key role in general. Some stated that the length of a learning video should be limited to a maximum of about 10 minutes. After that, it becomes more difficult to maintain attention and absorb the learning content meaningfully.

---

<sup>5</sup><https://imoox.at/mooc/> (last access November 2024)

<sup>6</sup><https://metacampus.unite-university.eu/> (last access November 2024)

### 3 Results

Advantages of AI in Learning Environments			
Category	Description	Example	Frequency
Overcoming Language Barriers	This category includes language challenges and preferences based on the use of an AI-generated avatar in foreign languages.	"I would definitely prefer voice translation because I'm not a big fan of subtitles." "The technology can be used to overcome language barriers."	38
Efficiency and Time saving for Video Creation	This category describes the practical benefits of AI avatars, such as the ability to produce videos more efficiently in terms of time, costs and expense.	"I can imagine that AI avatars will be used more frequently in the future, especially in areas where a lot of information needs to be conveyed. But I don't think they can completely replace the human teacher."	28
Fast and Easy Content Generation	This category includes how participants are already using artificial intelligence to generate content, access information, generate ideas and consume basic knowledge.	"I find the use of AI, such as ChatGPT for idea generation or text finding, very helpful."	20
Equal Educational Opportunities	This category refers to how AI technology can impact equal opportunities in education in terms of time, location and financial independence.	It would not be possible to offer education across the board, regardless of location, time or social reputation.	9

Table 3.8: Inductive category table which potentials and advantages can be gained from using AI in the context of learning environments with corresponding sub-categories.

By far the most frequently mentioned benefits of using AI in learning environments, shown in Table 3.8, were overcoming language barriers and making it easier to create videos in terms of time and cost on the teaching side. Overcoming language barriers primarily meant automatic translation using AI into the respective native language, which would lead to an increase in the amount of learning content consumed and would also allow this to take place regardless of location and in a socially just manner. This statement is also clearly reflected in the personal attitude towards linguistic translations using AI instead of transcripts, illustrated in Table 3.11.

### 3 Results

When it comes to efficiency and time factors concerning creating learning videos, the majority agreed on strong advantages of using AI in this specific context. Teachers would not have to spend huge amounts of time recording learning videos or getting to the video studio. This would also lead to savings on expensive equipment. In case of missing on-screen experience of instructors, the use of AI was considered beneficial too and could offer invaluable value in the form of technical support. In addition, some participants mentioned the adaptation of existing video material as an advantage of using AI in learning environments. The use of AI could even lead to an improvement in quality, as learning material in videos can be edited or supplemented without having to record the video again, which can even have a positive effect on the sustainability of producing digital content.

Disadvantages of AI in Learning Environments			
Category	Description	Example	Frequency
Potential for Cheating or Plagiarism	This category portrays the potential of generated content in terms of irresponsible use.	"One can certainly expect an increased potential for cheating or plagiarism, for example in text generation for academic papers."	5
Losing the Learning Process	This category describes that some participants found that using AI can lead to loss in the learning process itself.	"When we rely on AI, the learning process is lost or laziness is encouraged."	4

Table 3.9: Inductive category table which disadvantages can be derived from using AI in the context of learning environments with corresponding sub-categories.

The most disadvantages mentioned by participants, were the potential for cheating through using AI for text generation in the context of learning. Also the learning process itself could suffer through encouraging laziness of students. Another problem is the lack of questioning of the generated content, as it is usually not questioned and often no sources are given.

### 3 Results

Threats about AI			
Category	Description	Example	Frequency
Credibility, Trust and Ethics	Some participants expressed skepticism about the use of AI avatars, particularly with regard to possible misuse or ethical concerns.	"I'm not sure I want AI to play such a big role in education. There are so many ethical and technical questions that are still open and I think we should approach it with caution."	11
Skepticism about usage of AI	This category includes reservations about the technology by some participants.	"Dangerous of taking on knowledge without questioning, influencing or thinking for yourself." "No, scary."	9
Privacy Concerns	Some participants expressed concerns of using AI regarding their privacy.	"Yes, but there are concerns about data protection e.g. deletion, misuse or even in case of death."	9
Fear of Job Losses	This category includes fears that the use of AI avatars could replace human teachers and lecturers in the future, which could endanger jobs and fundamental experience.	"I wonder what this means for teachers and lecturers. If AI avatars are able to teach learning material just as well, then real people could eventually lose their jobs."	2

Table 3.10: Inductive category table which threats emerged from using AI in the context of learning environments with corresponding sub-categories.

Table 3.10 illustrates the biggest fears about using AI in learning environments, mentioned by the participants. By far the skepticism about the usage of AI in this context as well questioning the credibility, trust and ethics play an important role. In connection with the disadvantages of AI (see Table 3.9), questioning AI generated content, whether text, audio or video, some participants lack trust in the usage of AI based on social media reports, deep fakes, or personal experiences. In addition, it is not clear what happens to the personal data provided and how the AI processes it in the background, which led some participants to express concerns about data protection. Only a few participants had concerns about the teacher being completely replaced by the AI itself.

### 3 Results

Personal Perspective			
Category	Description	Example	Frequency
Preference for Translation using AI	Participants commented on the possibility of using AI to translate learning videos into native languages if the quality is sufficient.	"Especially when it comes to learning videos, I think it's cool when they are automatically translated into your own language."	27
Technological Curiosity	This category summarizes the interest and curiosity in the technology such as creating an own AI avatar.	"It would certainly be an interesting experience to have an AI avatar of myself." "I would like to try it out."	16
Preference for real Instructor	Some participants stated that they prefer to see real people in learning videos because they find this more authentic and engaging.	"I would rather see a real person in the video because the body language is different, there isn't that monotony, the sameness, and yet it's still something different and comes across differently."	9
Awareness	This category describes the importance of awareness using AI in the educational and private context.	"Creating awareness to use AI not only negatively, but also as a practical tool." / "Knowing how prompting works, for example, is also important."	3

Table 3.11: Inductive category table which personal opinions and perspectives can be derived from using AI in the context of learning environments with corresponding sub-categories.

Personal preferences of the participants were assigned in Table 3.11. In connection with the advantages of AI in learning environments in Table 3.8, the majority prefers linguistic translation with image and sound of educational content into the native language by AI techniques instead of reading translated transcripts. Also a few participants prefer real instructors instead of AI generated instructors based on sympathy, authenticity and engagement, which refers strongly to the *Emotional Learning* aspect described by the category of *Learning Experience* illustrated in Table 3.7.

### 3 Results

Nevertheless, the majority of participants stated that the quality of AI-generated avatars for learning videos was sufficient and therefore would not make a significant difference compared to real lecturers as illustrated in Figure 3.14.

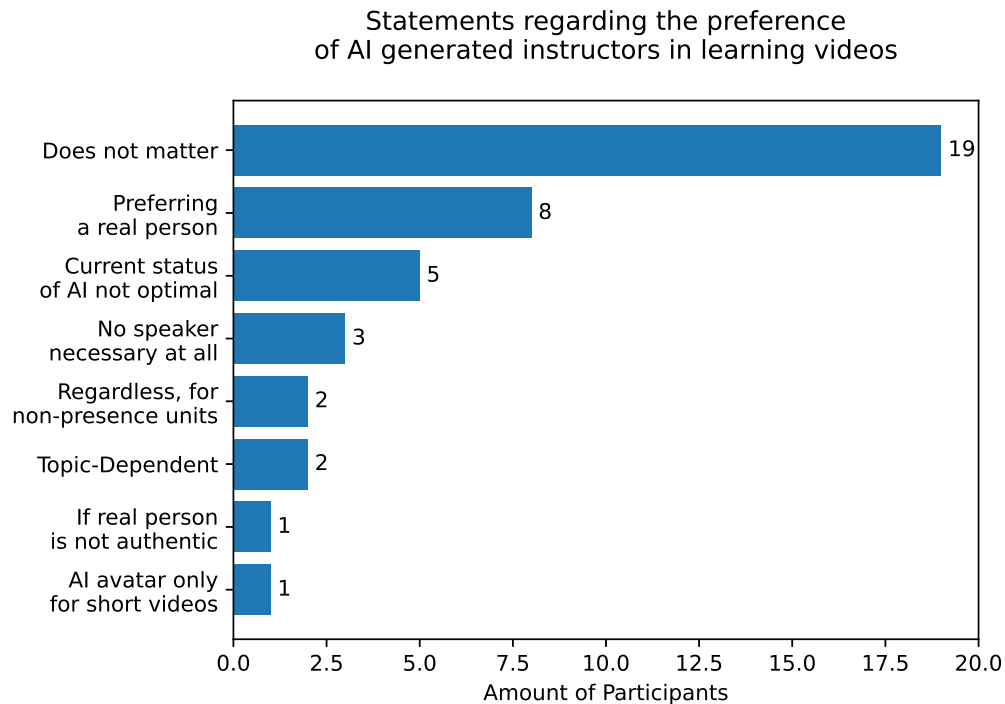


Figure 3.14: Preference of participants using AI generated instructors in learning videos.

The majority of about 46% participants of the interviews (n=41) would create an AI avatar by themselves (see Figure 3.15) which indicates a strong curiosity about the technology itself. Nevertheless, most participants would only have an AI avatar created in conjunction with a learning video of themselves. Most of them would also not make a personal AI avatar available on public platforms. This clearly raises concerns about data protection which can also be obtained in Table 3.10, illustrating most common threats regarding the usage of AI. Following on from this point, the fate of the AI avatar after the death of a person was also frequently discussed.

### 3 Results

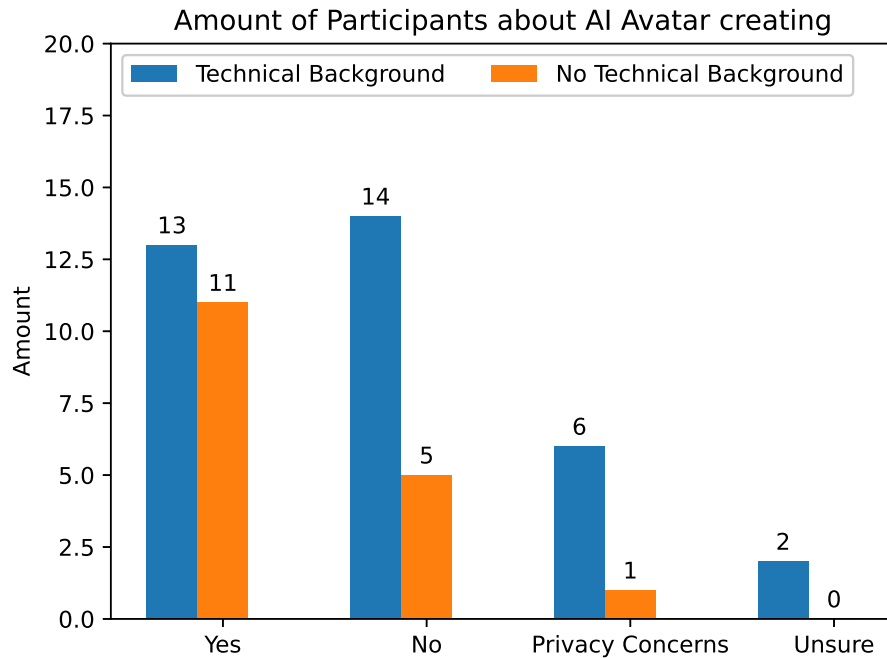


Figure 3.15: Participants who would generate an AI avatar of themselves grouped by their educational background.

Another aspect of the analysis was the willingness of creating an AI avatar of the individual participants covered in the face-to-face interview after the micro courses. The technical background refers to former experience with AI tools in general regardless of whether by using them for audio, text or video generation. No technical background refers to poor prior knowledge or simply heard about AI in the media. Two people did not answered in a clear way and one transcript could not be obtained fully during the transcript process with Whisper. Therefore the analysis ended up with (n=52) answers in total grouped by the corresponding educational background of each participant. The graphic 3.15 clearly shows that participants with prior knowledge concerning AI techniques, tend not to have an avatar created of themselves compared to participants with no prior knowledge who would create an AI avatar of themselves more likely.



### 3 Results

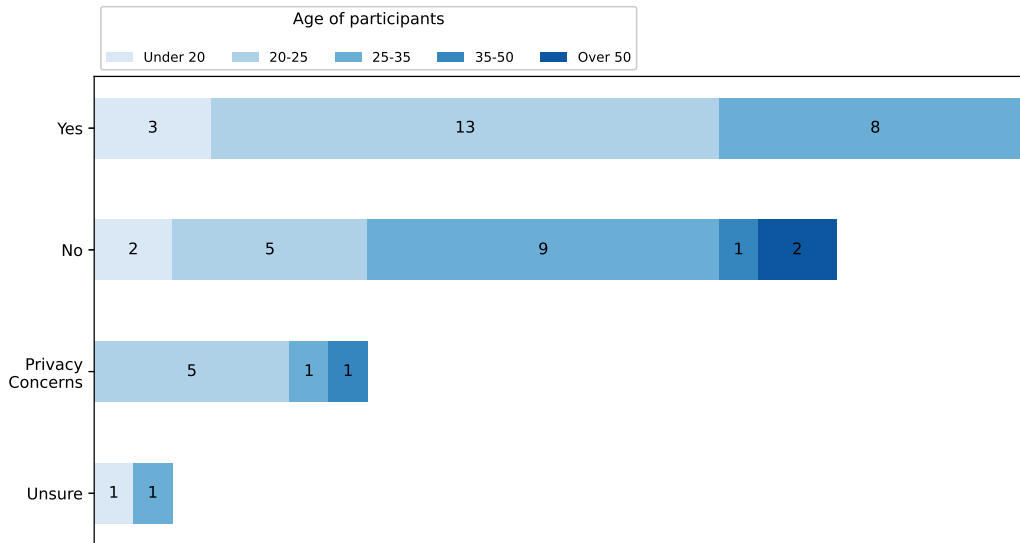


Figure 3.16: Participants who would generate an AI avatar of themselves grouped by their age.

The curiosity of creating an AI avatar of participants for themselves grouped by age ended up the same way as in 3.15 with (n=52) participants in total for the same reasons. The illustration 3.16 shows the age distribution over all valid participant statements about AI avatar creating.

Participants in the range of 20-25 years tend to create an avatar for themselves more likely while participants in the range of 25-35 seem to be conflicted about it. With a total amount of four participants older than 35 years (n=4), the usage of own AI avatars is rejected completely or due to privacy concerns. Also privacy concerns only emerge between the age of 20-50 years, excluding younger and older people.

### 3 Results

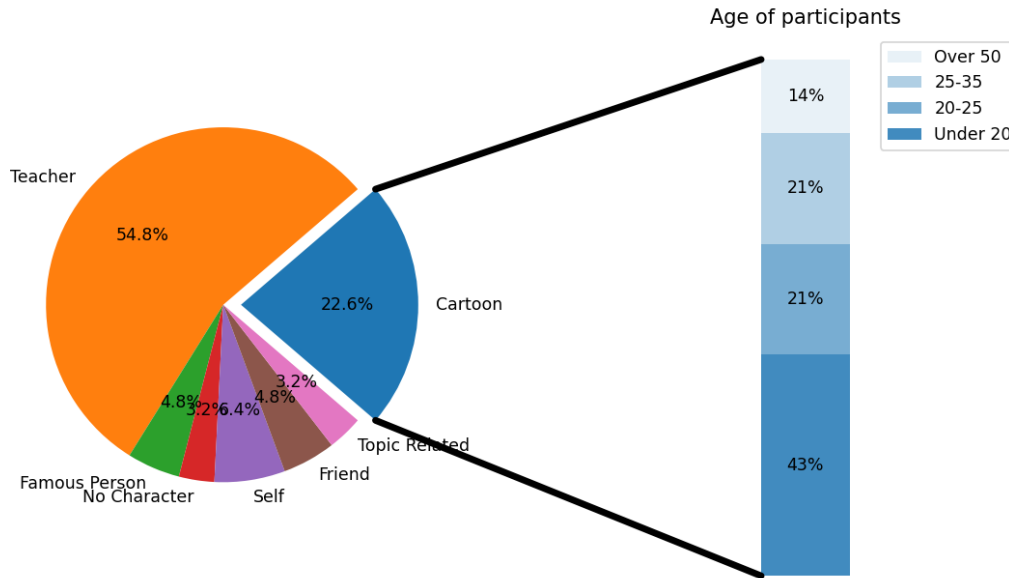


Figure 3.17: AI Avatar preferences of participants

Considering the preferences for the type of AI generated avatars in learning environments in higher education, the majority (58.4%) of participants would prefer a teaching person for the knowledge transfer. The most frequently cited reasons for the outcome were seriousness, credibility and authority.

The second choice would be cartoon-like instructors (22.6%) for a playful transfer of knowledge and to characterize artificial intelligence as such with a certain degree of alienation. A recently published study by Pataranutaporn et al. (2022) deals with usage of generative AI for customization of video instructors and their effects on improvement of students' motivation towards learning in more detail as well.

The minority in descending order would choose themselves with an average participants' age of under 20 (6.4%), generated friends, famous persons, idols, former experts (4.8%) or even no character at all (3.2%) in learning videos. A few participants (3.2%) would select the type of the generated instructor based on the learning content.

### 3 Results

Future of AI in Learning Environments			
Category	Description	Example	Frequency
Desired Future Use	Many participants expressed positive views on the future use of AI in learning videos. There is also great potential for supporting measures.	"Yes, because I think it will perhaps bring improvements in the future in principle I think it can be quite cool and also make work easier."	32
Quality Assurance	Some participants express concerns about whether they can trust the content delivered by an AI avatar as much as they would a human presenter.	"With a human, you know that there is experience and knowledge behind it, but with a machine it's somehow different." "Text must be approved by experts."	11
Labeling in Learning Environments	Some believe that the use of AI in learning environments should be labeled as such while others think this can lead to just focusing on the AI instead.	"There are a lot of concerns when you don't know if the person is genuine. Trust is important, especially in learning environments."	9
Individual Learning	This category refers to how the avatar could theoretically convey personalized content as the technology continues to develop.	If AI could eventually respond to my learning needs, that would be a real step forward. Maybe it could adapt to my learning pace or interests."	9

Table 3.12: Inductive category table considering the future use and necessary framework conditions in the context of learning environments with corresponding sub-categories.

The majority of participants agreed on using AI techniques in learning environments as the further development of technology can bring many advantages in this area in the future. In order to be able to use this technology effectively, certain framework conditions regarding quality assurance by experts and appropriate labeling of AI technology must be met.

### 3 Results

In addition to the qualitative content analysis regarding face-to-face interviews (n=41), a workshop was also held with a graduating class of The Federal Upper Secondary School Monsberger Graz<sup>7</sup> on the topic of artificial intelligence in learning environments. Through the absence of three people after the video session with the micro courses, the analyses ended up with a total of (n=11) participants aged 17-21.

During the workshop, the statements made by the participants on pre-defined categories were recorded in the form of a work using posters for each category respectively illustrated in Tables 3.13, 3.14 and 3.15.

#### Perception

Perception of the AI Avatar		
Statement	Positive	Negative
Slow perception of language and movement of the AI generated avatar		x
The emphasis of words was not always appropriate or was generally very monotonous		x
The gestures seemed blunt and inhuman		x

Table 3.13: Statements of students during the workshop regarding first impressions of the AI avatar.

Although the perception of the AI generated avatar seemed to be focused on abnormalities illustrated in Table 3.13, the majority of this group (more than 60%) did not recognize the AI avatar during watching the learning videos and appeared neutral or authentic to the same number of students.

---

<sup>7</sup><https://www.borg1.at/> (last access September 2024)

### 3 Results

#### Learning Experience

Learning Experience with the AI Avatar		
Statement	Positive	Negative
The use of AI in learning environments can support work.	x	
The use of AI-generated avatars makes no difference in learning compared to a real person presenting in a learning video.	x	

"I can understand all information equally well with AI avatars"	
<i>Yes</i>	6
<i>Neutral</i>	4
<i>No</i>	1

"I prefer to see a real instructor in learning videos"	
<i>Yes</i>	1
<i>Neutral</i>	9
<i>No</i>	1

"I would have a learning video translated with ..."	
Text-based with Subtitles	5
Generated voice and lip syncing by AI	6

Table 3.14: Statements of students during the workshop regarding learning experience with the AI avatar with assessments concerning comprehensibility, instructor preference and translation preference in learning videos.

As illustrated in Table 3.14 the overall learning experience with AI generated instructors make no difference compared to real instructors in learning videos for the majority of participants in the secondary school.

### 3 Results

#### Future of AI in Learning Environments

Future of AI in Learning Environments		
Statement	Positive	Negative
This allows information to be processed and made available much faster	x	
The acceptance of important tasks can be handed over to the AI, which could influence the learning process.		x
Sources are not specified or cannot be verified		x
Relying too much on AI		x
AI is constantly available for feedback	x	

This is where I want to use AI in the future	
Work	9
University	8
School	8
Leisure Time	6

Table 3.15: Statements of students during the workshop regarding the future use of AI in learning environments and possible advantages and disadvantages.

According to results shown in Table 3.15, participants stated that AI can make information dissemination faster and AI can be accessed at any time for possible questions, which is one of the major advantages in learning environments. However, some participants fear that by using AI support, important tasks that should be done by the learner will be handed over to the AI, thus losing the learning process. Blind reliance on AI could also lead to misinformation instead of learning success which was also mentioned by other participants in the face-to-face interviews.

## 3.4 Stimulus Video

The second research question of this thesis deals with the validation of outcomes considering the usage of the online face-reader software application FaceReader Online for future similar experimental setups. In order to provide a proper answer, participants' emotional responses during the video sessions were scored and compared to the personal insights and opinions during the face-to-face problem-centered interview after the emotional classification.

For this purpose the final stimulus video after the two micro-courses, one experimental sequence with generated instructors and one control sequence with real instructors, was shown to all participants with a comparison of the real instructor and the corresponding virtual instructor, illustrated in Figure 3.18, to reveal the study content in the end. Since the total amount of participants conducted the micro-courses with alternating instructors, more precisely with male and female instructors to reduce biases, two final videos were created solely showing the comparison of the two corresponding instructors - either male or female.



Figure 3.18: Comparison of the real female and virtual instructor at the end of the learning session.

### 3 Results

In order to provide a proper evaluation of the facial analysis tool working with emotional responses and their classification, an emotional indicator must be made available. Therefore the final stimulus video of the survey deals with the surprise effect and more precisely with the resolution and recognition of the AI generated avatar. Figure 3.12 illustrates the distribution of (N=55) participants recognizing or not recognizing the AI generated avatar, extracted during the face-to-face interview.

Since almost two-thirds (58.2%) of all participants did not recognize that one of the micro-courses was presented by an AI generated avatar, one of the most strongest emotions had to be “**Surprised**”, which served as one of the significant emotional indicators for the evaluation.

For the sake of analysis, grouped result over all participants were used instead of individual emotional responses with basic expressions on a timeline, illustrated in Figures 3.19 and 3.20 to analyse to whole emotional journey the final stimulus video provoked.

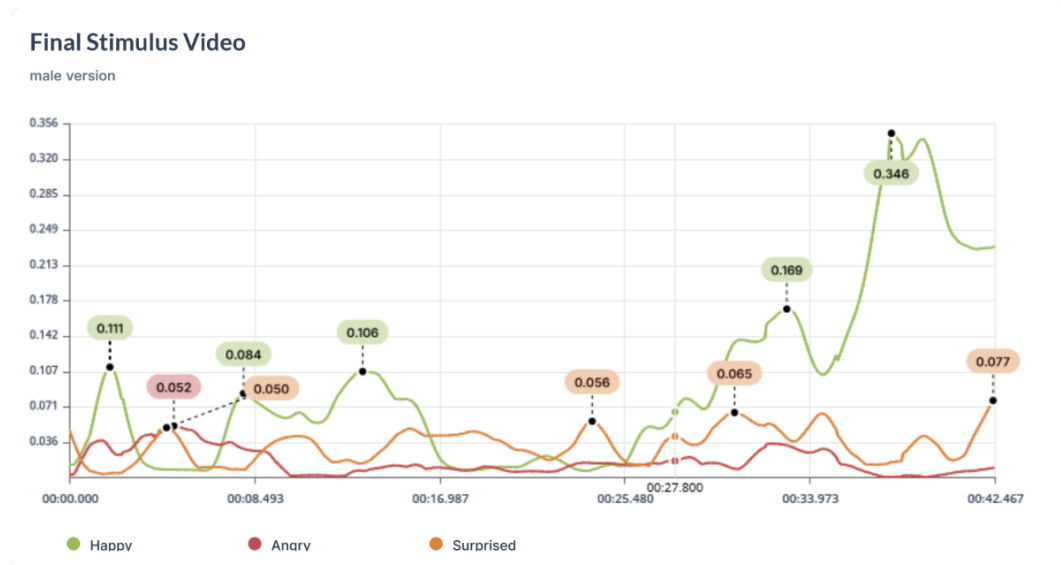


Figure 3.19: Participants emotional journey during the final stimulus video comparing the real male and the virtual male instructor with the strongest basic emotional expressions.



### 3 Results

The timeline chart in Figure 3.19 illustrates the three strongest captured emotions and their respective values (see Table 3.16) during watching the final stimulus video, namely **Happy**, **Angry** and **Surprised**.

Happy		Angry		Surprised	
Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation
0.075	0.0885	0.03	0.058	0.0195	0.035

Table 3.16: Significant and strongest captured emotions during watching the final stimulus video of the male versions.

The revealing sequence with the breakdown of the study topic happened at around second 8 where an increase in happy and slightly afterwards in surprised emotions can be obtained clearly. At around second 35 another control scene, serving for the evaluation of emotional classification, was added with the assumption of being funny. The face-reader software recognized a strong increase in happy emotions clearly as well. Striking peaks for surprised facial responses could be obtained throughout the whole stimulus video.

### 3 Results

Analogous to the male version, the timeline chart illustrated in Figure 3.20 containing the strongest captured emotions and their respective values (see Table 3.17) during watching the female version of the final stimulus video.

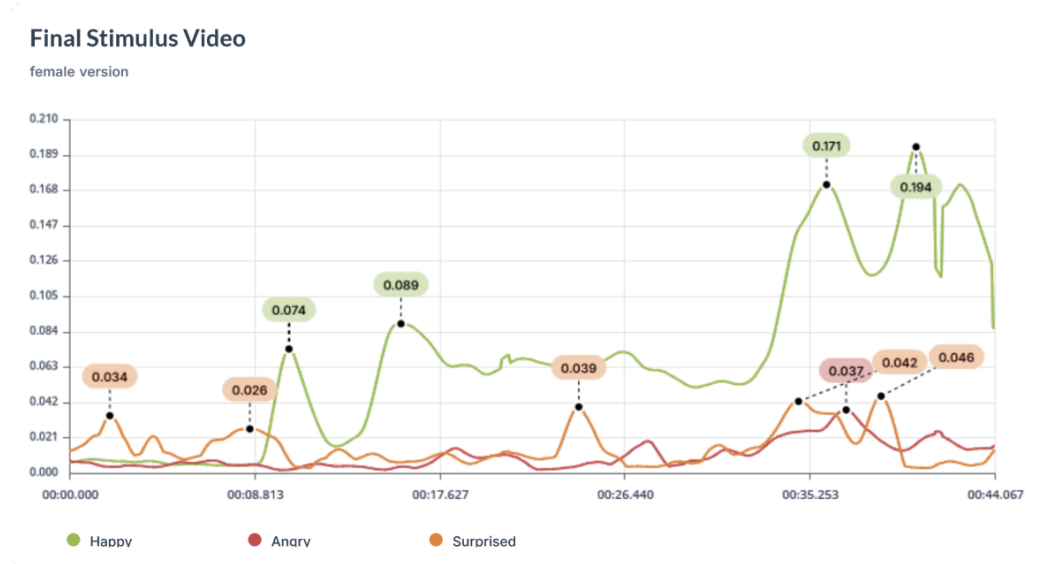


Figure 3.20: Participants emotional journey during the final stimulus video comparing the real female and the virtual female instructor with the strongest basic emotional expressions.

Happy		Angry		Surprised	
Average	Standard Deviation	Average	Standard Deviation	Average	Standard Deviation
0.1415	0.215	0.0155	0.0325	0.015	0.0245

Table 3.17: Significant and strongest captured emotions during watching the final stimulus video of the female versions.

Starting from our control point at around second 8, happily facial responses could be captured by the facial analysis tool clearly. In contrast to the male version of the stimulus video, the female version seemed to evoke more joyful emotions than surprised emotions among the participants. The second stimulus video still recognized striking peaks for surprised facial responses throughout the whole watching time and a clear increase of happy emotions at the second control sequence at around second 35.

## 4 Discussion

The objective of the underlying study of this thesis is to investigate and determine significant differences in the learning behavior and outcome of participants between AI-generated presenters as opposed to real presenters in learning videos in higher education. Particular attention was paid to the learning engagement of the participants with regard to differences in comprehensibility, attention and knowledge acquisition and more precisely in the evaluation of emotional and cognitive engagement. The aim was to check whether the current state of artificial intelligence can already be used sensibly for learning environments with AI-generated lecturers in learning videos without the learning process suffering.

To answer the predefined research questions (see chapter 1) the analysis of the study was divided into a quantitative and qualitative part and evaluated accordingly. The quantitative analysis includes the evaluation of post-assessments (see 3.1), the classification of emotions in the course of the facial analysis during the learning videos (see 3.2) as well as essential and clearly quantifiable components from the subsequent face-to-face interviews (see 3.3) including the workshop conducted. The qualitative analysis includes the evaluation of frequently made statements on the use of AI in learning environments during face-to-face interviews 3.3 using the methodology of the qualitative content analysis according to Mayring and Fenzl, 2019.

The first research question deals with the usage of AI instructors in learning videos and how do they affect their quality from the learners' perspective in higher education. From the quantitative perspective of the evaluation of the post-assessment, more precisely the subsequent single-choice questions related to the fact-based or content-related context of the previously shown learning video, the evaluation was in favor of the artificially generated

## 4 Discussion

lecturers. More questions were answered correctly in a learning video presented by an AI avatar than in learning videos with a real presenter as illustrated in Figure 3.1.

The classification of the participants' emotional standpoint, which was carried out using the facial analysis software FaceReader Online<sup>1</sup>, showed only minimal emotional distinctions while watching the learning videos, with the majority indicating a neutral emotional mood (see Figure 3.3). The arousal-valence model, illustrated in Figure 3.11, also clearly confirmed the minimal difference in the participants' emotional engagement with regard to artificially generated speakers compared to real speakers in learning videos. Due to the participants' references to subsequent questions on a topic, a continuous attention of the face-reader software could be determined in 100% of the participants (see Figure 3.9), which also clearly underlines the findings of Al-Shabandar et al., 2018 regarding incentives for learning with educational videos when undertaking a task.

The face-to-face interviews related to the first research question show clearly quantifiable results as well as subjective impressions of the participants. On the one hand, the clear majority could not identify the artificially generated speaker as such or notice any difference to a real speaker in a learning video, despite the speaker's familiarity and technical background on the subject of AI. Furthermore, the majority also stated that the use of artificially generated avatars in learning videos was irrelevant or not decisive for the absorption of learning content in contrast to real speakers in learning videos.

On the other hand, the qualitative content analysis clearly shows that the AI lecturer evoked an unnatural perception, which could not be assessed by most participants. It was also found that some participants paid less attention to the learning content, as they focused more on such inconsistencies and gestures, facial expressions and emphasis. The qualitative analysis also showed a general skepticism towards the use of AI in learning environments. For many of the respondents, labeling of artificially generated content, quality assurance or cross-checking of the educational content as well as the use of personal data for compliant and large-scale use of AI in

---

<sup>1</sup><https://facereader-online.com/> (last access November 2024)

## 4 Discussion

learning environments is a major concern. In contrast, however, some advantages have been identified through the use of AI in learning environments, such as overcoming language barriers, increasing the quality of learning through post-processing of learning content, minimizing time and financial expenditure for large-scale provision of knowledge and the positive use for individual learning.

The second research question deals with the evaluation of the facial analysis software FaceReader Online for a meaningful recognition of reliable emotional states in general. For this purpose, a stimulus video specially created for this evaluation was used at the end of the learning videos. The video contains certain sequences in which previously defined emotions were to be demonstrated. The software was able to clearly classify the prior defined emotions in the desired sequences, thereby confirming the appropriate and successful use of this type of software for emotional categorization as explained in more detail in section 3.4. Despite a satisfactory classification of emotions, some negative aspects were nevertheless noticed during the use of the software:

- Appearance of unusually long loading times in terms of evaluating group results
- Limited options for comparing individual tasks per participant
- Export of analytics data took several hours (N=55)
- Data analysis sometimes aborted with an undefined error message
- Participants with glasses or squinting participants affected the recording quality significantly

The third research questions was also answered by the classification of emotional states of the participants during watching the learning videos. This question refers to one of the three main aspects for classifying learning engagement, more precisely *Emotional Engagement* (see section 1.3.2), which has already been discussed in more detail by Struger, Br  nner, and Ebner, 2024. As already mentioned, no significant difference between artificially generated speakers and real speakers in learning videos could be found in the course of emotional classification during the consumption of the learning videos showed in section 3.2.

## 5 Conclusion

The current development of artificial intelligence in the field of knowledge transfer in learning environments has great potential in terms of accessibility, generation of qualitative learning material and learning engagement on a content and emotional level.

From an objective and quantitative point of view in the course of this thesis and its underlying empirical study, no demonstrably detrimental difference could be found between artificially generated lecturers and real lecturers in learning videos. However, unmeasurable and subjective perceptions of learners must not be ignored. Direct feedback makes it clear that with regard to the use of AI in learning environments, a large proportion of participants are skeptical about the reliability, trustworthiness and guarantee of the accuracy of the content. The spread of malicious use of AI and deep fakes outside the learning context also plays a major role and influences the perception of AI in general. For this purpose, uniform frameworks and conditions must be created in terms of transparent quality control, type of use and data compliance. The development of digital inclusion and individual learning is also playing an increasing role and can provide targeted support in the learning context with the help of AI such as *Adaptive Learning Platforms* or *Intelligent Learning Management Systems (ILMS)*.

In terms of learning engagement, the implementation of emotional attachment in terms of signaling, gestures and facial expressions of artificially generated speakers is also required to convey a more interactive and emotional overall picture in the future. Further research in this field must be constantly pushed forward and should always have a qualitative component with direct feedback in addition to quantitative aspects, since subjective feelings and personal viewpoints or concerns when using and dealing with AI cannot be measured effectively or can be difficult to represent.

## 6 Outline

There is no denying that the current status of AI development in the field of learning environments already offers a great impact and numerous benefits. However, further development in this field will still have to deal with some adjustments for a meaningful and equal approach in contrast to manual preparation of high-quality learning content. Building equal emotional connection for learning engagement with clear indications of the use of AI as well as legal and social frameworks remains a major challenge and is overshadowed by current misuses of AI. In the meantime, there are also the first regulations issued, such as the EU's Artificial Intelligence Act<sup>1</sup>, which also regulates the way of using an Emotion Recognition System (ERS) and came into force between the conduct of the experiment, evaluation of the data and submission of this thesis.

The progressive development and usage of AI driven real-time analytic tools have the potential of getting instant results to respond immediately to changes in learners' emotional impressions, e.g. for the use in *Adaptive Learning Systems (ALS)* or *ILMS*. Tools like the desktop version of FaceReader<sup>2</sup> by Noldus (in contrast to a necessary prefabricated experiment setup in the online version) already have the potential to evaluate learners' engagement during the learning process and to actively shape students' learning experience based on proven facial expression analysis technology. In combination with further developments of innovative video platforms like HeyGen<sup>3</sup> using the power of generative AI to create video content visually and auditory will certainly bring about major changes in learning environments and interactive learning methods which are based on more standardized specifications and legal compliance in the future.

---

<sup>1</sup><https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

<sup>2</sup><https://www.noldus.com/facereader> (last access November 2024)

<sup>3</sup><https://www.heygen.com/> (last access November 2024)

## Bibliography

- Beege, Maik, Manuel Ninaus, et al. (2020). "Investigating the effects of beat and deictic gestures of a lecturer in educational videos." In: *Computers & Education* 156, p. 103955. ISSN: 0360-1315. DOI: <https://doi.org/10.1016/j.compedu.2020.103955>. URL: <https://www.sciencedirect.com/science/article/pii/S0360131520301536> (cit. on p. 10).
- Beege, Maik, Sascha Schneider, et al. (2017). "Look into my eyes! Exploring the effect of addressing in educational videos." In: *Learning and Instruction* 49, pp. 113–120. ISSN: 0959-4752. DOI: <https://doi.org/10.1016/j.learninstruc.2017.01.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0959475217300270> (cit. on p. 10).
- Beege, Maik, Noah L. Schroeder, et al. (2023). "The instructor presence effect and its moderators in instructional video: A series of meta-analyses." In: *Educational Research Review* 41, p. 100564. ISSN: 1747-938X. DOI: <https://doi.org/10.1016/j.edurev.2023.100564>. URL: <https://www.sciencedirect.com/science/article/pii/S1747938X2300057X> (cit. on p. 11).
- Brame, Cynthia J. (2016). "Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content." In: *CBE—Life Sciences Education* 15.4. PMID: 27789532, es6. DOI: [10.1187/cbe.16-03-0125](https://doi.org/10.1187/cbe.16-03-0125). URL: <https://doi.org/10.1187/cbe.16-03-0125> (cit. on p. 6).
- Elamir, Mona, Walid Al-Atabany, and Mohamed Eldosoky (Apr. 2019). "EMOTION RECOGNITION VIA DETRENDED FLUCTUATION ANALYSIS AND FRACTAL DIMENSIONS." In: (cit. on pp. 40, 42).
- Guo, Philip, Juho Kim, and Rob Rubin (Mar. 2014). "How video production affects student engagement: An empirical study of MOOC videos." In: pp. 41–50. DOI: [10.1145/2556325.2566239](https://doi.org/10.1145/2556325.2566239) (cit. on p. 7).



## Bibliography

- Ingavélez-Guerra, Paola et al. (2022). "Automatic Adaptation of Open Educational Resources: An Approach From a Multilevel Methodology Based on Students' Preferences, Educational Special Needs, Artificial Intelligence and Accessibility Metadata." In: *IEEE Access* 10, pp. 9703–9716. DOI: [10.1109/ACCESS.2021.3139537](https://doi.org/10.1109/ACCESS.2021.3139537) (cit. on p. 17).
- Leiker, Daniel et al. (2023). *Generative AI for learning: Investigating the potential of synthetic learning videos*. DOI: [10.48550/arXiv.2304.03784](https://doi.org/10.48550/arXiv.2304.03784). eprint: [2304.03784](https://arxiv.org/abs/2304.03784) (cit. on pp. 1, 13, 18).
- Marta-Lazo, Carmen, Sara Osuna-Acedo, and Javier Gil-Quintana (2019). "sMOOC: A pedagogical model for social inclusion." In: *Heliyon* 5.3, e01326. ISSN: 2405-8440. DOI: <https://doi.org/10.1016/j.heliyon.2019.e01326>. URL: <https://www.sciencedirect.com/science/article/pii/S2405844018365253> (cit. on p. 5).
- Martinez-Garcia, Aitor, Patricia Horrach-Rosselló, and Carles Mulet-Forteza (2023). "Evolution and current state of research into E-learning." In: *Heliyon* 9.10, e21016. ISSN: 2405-8440. DOI: <https://doi.org/10.1016/j.heliyon.2023.e21016>. URL: <https://www.sciencedirect.com/science/article/pii/S2405844023082245> (cit. on p. 1).
- Mayer, Richard (Dec. 2008). "Applying the Science of Learning: Evidence-Based Principles for the Design of Multimedia Instruction." In: *The American psychologist* 63, pp. 760–9. DOI: [10.1037/0003-066X.63.8.760](https://doi.org/10.1037/0003-066X.63.8.760) (cit. on p. 7).
- Mayer, Richard (Jan. 2009). *Multimedia Learning: Second Edition*. DOI: [10.1017/CB09780511811678](https://doi.org/10.1017/CB09780511811678) (cit. on p. 11).
- Mayring, Philipp and Thomas Fenzl (2019). "Qualitative Inhaltsanalyse." In: *Handbuch Methoden der empirischen Sozialforschung*. Ed. by Nina Baur and Joerg Blasius. Wiesbaden: Springer Fachmedien Wiesbaden, pp. 633–648. ISBN: 978-3-658-21308-4. DOI: [10.1007/978-3-658-21308-4\\_42](https://doi.org/10.1007/978-3-658-21308-4_42). URL: [https://doi.org/10.1007/978-3-658-21308-4\\_42](https://doi.org/10.1007/978-3-658-21308-4_42) (cit. on pp. 25, 28, 29, 44, 66).
- Moorhouse, Benjamin Luke, Marie Alina Yeo, and Yuwei Wan (2023). "Generative AI tools and assessment: Guidelines of the world's top-ranking universities." In: *Computers and Education Open* 5, p. 100151. ISSN: 2666-5573. DOI: <https://doi.org/10.1016/j.caeo.2023.100151>. URL: <https://www.sciencedirect.com/science/article/pii/S2666557323000290> (cit. on p. 12).

## Bibliography


- Murtaza, Mir et al. (2022). "AI-Based Personalized E-Learning Systems: Issues, Challenges, and Solutions." In: *IEEE Access* 10, pp. 81323–81342. DOI: [10.1109/ACCESS.2022.3193938](https://doi.org/10.1109/ACCESS.2022.3193938) (cit. on p. 17).
- Nass, Clifford, Jonathan Steuer, and Ellen Siminoff (Jan. 1994). "Computer are social actors." In: p. 204. DOI: [10.1145/259963.260288](https://doi.org/10.1145/259963.260288) (cit. on p. 9).
- Ng, Yen Ying and Adam Przybyłek (2021). "Instructor Presence in Video Lectures: Preliminary Findings From an Online Experiment." In: *IEEE Access* 9, pp. 36485–36499. DOI: [10.1109/ACCESS.2021.3058735](https://doi.org/10.1109/ACCESS.2021.3058735) (cit. on pp. 8, 10, 11).
- Park, S. (Jan. 2015). "The effects of social cue principles on cognitive load, situational interest, motivation, and achievement in pedagogical agent multimedia learning." In: 18, pp. 211–229 (cit. on p. 9).
- Pataranutaporn, Pat et al. (2022). "AI-Generated Virtual Instructors Based on Liked or Admired People Can Improve Motivation and Foster Positive Emotions for Learning." In: *2022 IEEE Frontiers in Education Conference (FIE)*, pp. 1–9. DOI: [10.1109/FIE56618.2022.9962478](https://doi.org/10.1109/FIE56618.2022.9962478) (cit. on pp. 1, 18, 57).
- Pellas, Nikolaos (2023). *The influence of sociodemographic factors on students' attitudes toward AI-generated video content creation*. DOI: <https://doi.org/10.1186/s40561-023-00276-4> (cit. on p. 12).
- Reeves, Byron and Clifford Nass (Jan. 1996). "The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Pla." In: *Bibliovault OAI Repository, the University of Chicago Press* (cit. on p. 9).
- Schacter, Daniel and Karl Szpunar (Feb. 2015). "Enhancing Attention and Memory During Video-Recorded Lectures." In: *Scholarship of Teaching and Learning in Psychology* 1. DOI: [10.1037/st10000011](https://doi.org/10.1037/st10000011) (cit. on p. 7).
- Al-Shabandar, Raghad et al. (2018). "Analyzing Learners Behavior in MOOCs: An Examination of Performance and Motivation Using a Data-Driven Approach." In: *IEEE Access* 6, pp. 73669–73685. DOI: [10.1109/ACCESS.2018.2876755](https://doi.org/10.1109/ACCESS.2018.2876755) (cit. on pp. 14, 67).
- Stephen Downes, George Siemens (Dec. 2008). *Connectivism and Connective Knowledge*. URL: <https://www.downes.ca/presentation/422> (cit. on p. 5).
- Struger, Patrick, Benedikt Brünner, and Martin Ebner (2024). "Promotion of Emotional Learning in Technical and Social Domains: A Systematic Review." In: *Learning and Collaboration Technologies*. Ed. by Panayiotis Za-

## Bibliography

- phiris and Andri Ioannou. Cham: Springer Nature Switzerland, pp. 241–255. ISBN: 978-3-031-61685-3 (cit. on pp. [15](#), [16](#), [68](#)).
- Van Vaerenbergh, Steven and Adrián Pérez-Suay (May 2022). “Intelligent Learning Management Systems: Overview and Application in Mathematics Education.” In: pp. 206–232. ISBN: 9781799892472. DOI: [10.4018/978-1-7998-9247-2.ch009](#) (cit. on pp. [3](#), [4](#)).
- Zhang, He et al. (2024). *Redefining Qualitative Analysis in the AI Era: Utilizing ChatGPT for Efficient Thematic Analysis*. DOI: [10.48550/arXiv.2309.10771](#). arXiv: [2309.10771](#) [cs.HC]. URL: <https://arxiv.org/abs/2309.10771> (cit. on p. [31](#)).

# Appendix

## A.1 Consent form for participants



Vizerektorat Lehre  
Lehr- und Lerntechnologien

We care about eEducation!

**Einwilligung zur Teilnahme an einem Versuch und zur Verwendung von Videoaufnahmen**

Hiermit erkläre ich, (im Folgenden „Teilnehmer:in“ genannt), freiwillig mein Einverständnis, an einem Versuch der Technischen Universität Graz (im Folgenden „TU Graz“) teilzunehmen. Im Rahmen dieses Versuchs erkläre ich mich damit einverstanden, dass Videoaufnahmen von mir während der Durchführung des Versuchs gemacht werden.

Diese Videoaufnahmen werden ausschließlich zu wissenschaftlichen Zwecken verwendet, insbesondere zur Auswertung der Wirkung eines Lernvideos auf mich als Teilnehmer:in. Die Auswertung der Videoaufnahmen erfolgt durch einen spezialisierten KI-Algorithmus der Firma Noldus Information Technology BV, Nieuwe Kanaal 5, 6709 PA Wageningen, Netherlands, auf Servern in Europa bereitgestellt durch Microsoft Azure.

**Zweck der Verarbeitung:**

Die erhobenen Videoaufnahmen dienen dem Zweck, die Effektivität und die Auswirkungen des Lernvideos auf die Teilnehmer:innen zu analysieren. Die Verarbeitung dieser Daten durch den KI-Algorithmus ermöglicht eine objektive Bewertung der Reaktionen und des Verhaltens der Teilnehmer:innen während des Versuchs.

**Datenschutz und Anonymität:**

Es wird garantiert, dass die Videoaufnahmen und die daraus resultierenden Daten anonym behandelt werden. Eine Zuordnung der Aufnahmen zu einzelnen Personen wird nicht vorgenommen. Die Videoaufnahmen werden nach Abschluss der Auswertung gelöscht.

**Betroffenenrechte:**

Ich bin darüber informiert, dass ich gemäß der Datenschutz-Grundverordnung (DSGVO) das Recht habe, Auskunft über die zu meiner Person gespeicherten Daten zu erhalten, die Berichtigung unrichtiger Daten zu verlangen, die Löschung oder Einschränkung der Verarbeitung meiner Daten zu fordern sowie Widerspruch gegen die Verarbeitung einzulegen. Weiterhin habe ich das Recht auf Datenübertragbarkeit. Für die Ausübung meiner Rechte oder bei Fragen zum Datenschutz kann ich mich an den Datenschutzbeauftragten der TU Graz wenden.

**Widerruf der Einwilligung:**

Mir ist bekannt, dass ich meine Einwilligung zur Teilnahme am Versuch und zur Verarbeitung meiner Daten jederzeit ohne Angabe von Gründen mit Wirkung für die Zukunft widerrufen kann. Der Widerruf berührt die Rechtmäßigkeit der aufgrund der Einwilligung bis zum Widerruf erfolgten Verarbeitung nicht.

**Schlussbestimmungen:**

Diese Einverständniserklärung wurde von mir freiwillig abgegeben. Ich wurde über den Zweck der Datenerhebung, -verarbeitung und -nutzung ausführlich informiert und habe die Möglichkeit, Fragen zu stellen, die zur vollständigen Klärung geführt haben.

Figure .1: First page of the consent form for participants for the recording and evaluation of biometric data with the FaceReader software as well as their legally compliant storage, processing and deletion.



Vor- und Nachname: \_\_\_\_\_

Ort, Datum: \_\_\_\_\_ Unterschrift des:der Teilnehmer:in: \_\_\_\_\_

Kontakte:

OE Lehr- und Lerntechnologien • Benedikt Brünner, MEd BEd • bruenner@tugraz.at

### Datenschutzinformation

Verantwortlicher:	TU Graz, Rechbauerstraße 12, 8010 Graz; Kontakt: <a href="mailto:datenschutz@tugraz.at">datenschutz@tugraz.at</a>
Datenschutzbeauftragter:	x-tention Informationstechnologie GmbH Römerstraße 80a, 4600 Wels Tel: +43 7242 2155 65065 Kontakt: <a href="mailto:datenschutzbeauftragter@tugraz.at">datenschutzbeauftragter@tugraz.at</a>
Daten:	Videoaufnahme, Foto, Name
Rechtsgrundlage:	<b>Einwilligung</b> gem. Art 6 Abs 1 lit a DSGVO
Empfänger:	Soweit es im Rahmen obiger Zweckbestimmung erforderlich ist, Dienstleister einzubeziehen, sorgt die TU Graz für die Einhaltung der datenschutzrechtlichen Bestimmungen. Davon abgesehen werden Foto- und Videoaufnahmen nicht an Dritte weitergegeben.
Drittlandsübermittlung:	Nein
Speicherdauer:	beschränkt auf den Zweck der Verarbeitung; bis zum Zeitpunkt des Widerrufs Ihrer Einwilligung
Betroffenenrechte:	<p>Sie haben ein Recht auf:</p> <ul style="list-style-type: none"><li>- <b>Auskunft</b>, um zu überprüfen, ob und welche personenbezogenen Daten wir über Sie gespeichert haben</li><li>- <b>Berichtigung/Vervollständigung</b> Ihrer personenbezogenen Daten, die falsch oder unvollständig sind</li><li>- <b>Löschung</b> Ihrer personenbezogenen Daten, die nicht (mehr) rechtskonform verarbeitet werden</li><li>- <b>Einschränkung der Verarbeitung</b></li><li>- <b>Datenübertragbarkeit</b></li><li>- <b>Freie Widerruflichkeit</b> einer erteilten Einwilligung mit Wirkung für die Zukunft; d.h. die Verarbeitung der davon betroffenen Daten wird – wenn nicht ein anderer Rechtfertigungsgrund vorliegt – ab diesem Zeitpunkt unzulässig</li><li>- <b>Recht auf Widerspruch</b>, soweit die TU Graz sich auf ein überwiegendes berechtigtes Interesse stützt (Ausnahme Medienprivileg § 9 DSG)</li></ul> <p>Bezüglich dieser Rechte finden Sie die Kontaktmöglichkeiten unter: <a href="https://datenschutz.tugraz.at/dsgvo/rechte/">https://datenschutz.tugraz.at/dsgvo/rechte/</a></p> <p>Sie haben auch ein <b>Beschwerderecht bei der Datenschutzbehörde</b>.</p>

Figure .2: Second page of the consent form for participants for the recording and evaluation of biometric data with the FaceReader software as well as their legally compliant storage, processing and deletion.

## A.2 Interview Questions

### Leitfaden für problemzentriertes Interview

Uhrzeit: \_\_\_\_\_ Personen ID: \_\_\_\_\_

#### Wahrnehmung zur KI-Software

**Leitfrage:** Hätten Sie sich gedacht, dass das erste Video mit einem KI Avatar erstellt worden ist?

*Unterfragen:*

- Wirkte der KI Avatar komisch oder unnatürlich auf Sie?
- Kennen Sie die Person im Video persönlich?
- Gab es eine besonders auffallende Stelle?
  - Falls ja, welche?
- Haben sie von dieser Technologie schon gehört?
  - Falls ja, woher oder in welchem Kontext?

#### Lernvideos mit KI-Software

**Leitfrage:** Können Sie mit dem KI Avatar gleich gut lernen und wenn ja/nein warum?

*Unterfragen:*

- Haben Sie eine Änderung in Bezug auf Wahrnehmung oder Auffassungsgabe zwischen den beiden Videos bei sich wahrgenommen?
  - Falls ja, welche?
- Konnten Sie sich im Video mit dem KI Avatar gleich gut Informationen merken bzw. inhaltlich verstehen?
- Möchten Sie lieber eine echte Person im Video sehen?
  - Warum bzw. warum nicht?

#### Zukunft von Lernvideos

**Leitfrage:** Möchten Sie, dass diese Technologie zukünftig eingesetzt werden wird und wenn ja warum bzw. wenn nein, warum nicht?

*Unterfragen:*

- Wenn Sie ein Lernvideo in einer anderen Sprache ansehen, wäre Ihnen ein Video in der originalen Sprache oder ein mit KI übersetztes Video lieber?
- Welchen Avatar würden Sie gerne in einem Lernvideo sehen? Sich selbst, den Lehrenden, einen Freund oder einen Cartoon Charakter?
- Welche Vorteile / Nachteile sehen Sie persönlich beim Einsatz von KI in Lernumgebungen?
- Würden sie gerne einen KI Avatar von sich selbst erstellen?

#### Daten für Personengruppenbeschreibung

- Studien Fachrichtung
- Alter
- Möchten Sie über die Ergebnisse der Studie informiert werden?
- Intern: Ist während dem FaceReader etwas vorgefallen?

Figure .3: Interview questions of the problem centered interview consisting of 3 topic blocks with 1 key question and corresponding sub-questions.

## A.3 Post Assessment Questions

### Frage 1 - MetaCampus

Welche Universität ist Teil von UNITE?

- TU Darmstadt (*RICHTIG*)
- TU Wien (*FALSCH*)
- Aalto Universität (*RICHTIG*)
- TU München (*FALSCH*)

### Frage 2 - MetaCampus

Auf welchem Open-Source Learning-Management-System basiert der Meta-Campus?

- Distance Education (*FALSCH*)
- Moodle (*RICHTIG*)
- UNITE! (*FALSCH*)
- Blackboard (*FALSCH*)

### Frage 1 - iMooX

Welchen Ansatz verfolgt iMooX?

- xMOOC Format (*RICHTIG*)
- cMOOC Format (*FALSCH*)
- openMOOC Format (*FALSCH*)

### Frage 2 - iMooX

Das „xMOOC“-Format bedeutet, dass Kurse hauptsächlich

- auf Video-Inhalten basieren (*RICHTIG*)
- den Lernfortschritt anzeigen (*FALSCH*)
- LIVE-Termine oder Webinare beinhalten (*FALSCH*)
- von Universitäten angeboten werden (*FALSCH*)



## A.4 FaceReader Online Result View



Figure .4: Interactive result view for participant #4 concerning the micro-course with the topic “MetaCampus” provided through the analytical backend of FaceReader Online.