Human Centered Computing in Digital Persona Generation

Nisha Ramachandra(✉), Manish Ahuja, Raghotham M. Rao, and Neville Dubash

Accenture Labs, Accenture, Bengalaru, India
{nisha.ramachandra,manish.a.ahuja,raghotham.m.rao, neville.dubash}@accenture.com

Abstract. Deepfake (or as we call it Digital Persona) has been very popularly used to create synthetic media in which a person in an existing image or video is replaced with someone else who is not present in that media. It refers to manipulated videos, or other digital representations produced by sophisticated artificial intelligence (AI), that yield fabricated images and sounds that appear to be real.

Deepfakes generally have been used for the purpose of defaming someone, where the user experience is not much of a concern. However, our work demonstrates using this technique for a good purpose. We created a digital persona of a renowned deceased artist with the aim to bring an enriching human experience through conversing with the persona projected on a 3d holographic stage in a museum. The digital persona responds in the voice of deceased artist to any questions asked by visitors related to his art journey and artwork. To ensure that the end results would have the audience immersed or awed with the outcome a.k.a. the digital persona, we adopted the human centered computing methodology which aims at radically changing the standard computing techniques of software development. In this work, the key elements of human centered computing include: a. Technology b. Cognitive Psychology and Ergonomics c. Social and Organizational Psychology d. Design and Arts e. Interaction f. Analysis for design of systems with a human focus from beginning to the end. We present the usage, details and outcomes of the mentioned focus areas in our design of developing deepfakes for good. We also present results of a social experiment conducted with children during their interaction with digital persona.

Keywords: Conversational deepfake · Human centered computing · Cognitive Psychology and Ergonomics · Social and Organizational Psychology · Design and Arts · Interaction · Analysis

1 Introduction

Deepfake means creating fake personas using deep learning methods. Deepfakes leverage powerful techniques from machine learning and artificial intelligence to

W. Fu et al. (Eds.): ICMTEL 2021, LNICST 388, pp. 385–401, 2021.
https://doi.org/10.1007/978-3-030-82565-2_32

manipulate or generate visual and audio content with a high potential to deceive [25].

A synthetic media is created by morphing the face. Speech synthesis techniques can also be applied to generate fake voice. Researchers have developed new software that uses machine learning to let users edit the text transcript of a video to add, delete or change the words coming right out of somebody's mouth [17]. Deepfakes have largely been used to create fake celebrity pornographic videos, fake news, hoaxes and financial fraud. In the above mentioned use cases, impersonation is the main objective and not the user centered design.

However, when used ethically, deepfakes can also be used for good. There can be many useful applications of deepfakes such as creation of visual storytelling in the field of education and films, creating real enough medical images to train AI in the medical field [23] and to originate a conversation to guide the end user towards positive thinking.

In such cases, the creator of the fake needs to consider the end user experience as a part of the design requirements. There is no existing framework for deepfake creation which can act as guide to generate an impressive deepfake within ethical limits.

We have applied the deepfake technology in the field of art to create a digital persona of a renowned deceased artist to help increase the awareness of Indian artworks. The goal of the digital persona creation is to enable learners during art and culture training sessions to better understand concepts from famous creators, curators and personalities.

The requirements of the digital persona were to interact with the users and create an effect of users having conversation with a real person by engaging the users in talks related to Indian art and art works through the life of the digital persona. To create this digital persona, we adapted the techniques of face superimposition [28] and speech synthesis [18] based on deep learning, natural language understanding for conversation and emotion detection [14] to measure the end user involvement during interactions. However, we believe that to create an "ideal" deepfake which is to be perceived as very realistic, it needs to be: a. immersive, b. expressive, c. responsive, d. proactive and e. adaptive; the definitions of which, used in our context are as below.

Immersive: There should be no visual flaw in creation of the deepfake for the end users to identify it as fake. The digital experience of the fake being projected on a screen should keep the end user engaged and focused.
Expressive: The digital persona should effectively and eloquently converse with the end user with appropriate expressions.
Responsive: The digital persona should correctly respond to the questions asked by the end user and reply diplomatically to any unfamiliar questions.
Proactive: If during the interaction, the end user deviates from the talks related to art works, the digital persona should proactively guide the end user to the art related topics.
Adaptive: Based on the involvement level and end user expression, the digital persona should understand the user mood and be flexible during the conversations.

To ensure adherence to the above listed characteristics, we conclude that, just the usage of the technology is insufficient for developing the digital persona and we propose to adapt the human centered computing design principles along with ethical guidelines.

This paper is structured as follows. In Sect. 2, we talk about the related work. In Sect. 3, we talk about the approach in depth by mentioning the implementation of each of the elements of human centered computing in detail. We discuss the outcome and social experiment results in Sects. 4 and 6 and conclude in Sect. 7.

2 Related Work

Generating good quality deepfakes is very well explored using deep learning techniques. The Dali museum in USA had already implemented digital persona of legendary artist Salvador Dali [6], however our approach goes beyond the Dali experience as our solution approach is conversational and employs human centric design principles. We also used speech generation using deep learning techniques and entire experience is not on a screen but is a 3D holographic projection. The basic architecture of all the deepfake generation techniques is same and uses the autoencoder with common encoder weights [15]. There are numerous enhancements over the basic deepfake architecture that people use for getting better results such as adding a discriminative network to the basic autoencoder architecture for training a generative discriminative network [16], where the decoder network of autoencoder acts as the generator network. The generator and discriminator network are then jointly trained to compete, resulting in better quality fakes. Other variations include learning with attention mechanism, adding multiple face extraction methods such as MTCNN and S3FD [12], adding perceptual loss [24,28].

Just like image synthesis, speech synthesis that can depict a person saying things is also very well explored. The speech synthesis algorithms use various parameters like intonation, tone, stress, rhythm, frequency spectrum etc. to generate voice from text. Deep learning is also used to train models using recorded speech data. Wavenet by Deepmind [30] and Tacotron by Google [36] are two such examples. While, multiple methods for creating good quality deepfakes through technology exists, to the best of our knowledge, there are no standard guidelines of using a well-defined framework/approach for creating the deepfakes with an immersive experience. There are no reference materials which talk about the various aspects of interface, interaction, design process and putting together the knowledge of people and technology in the creation of deepfake a.k.a. the digital persona. Based on our work, we have presented the design approach implemented, which we believe, is a well defined methodology and could be the baseline for any future works.

3 Approach

While deepfake technology can and has been used for unethical and illegal purposes, it can also be used ethically for social good too in the sample use cases mentioned in the Introduction section. For implementation in the social good scenario, it is important to consider the aspects of the deployment environment and study the user interactions during deployment.

The proposed approach focuses on leveraging the human centered computing design [20] to create a realistic digital persona using deepfake technology along with adherence to compliance and ethics. We believe that an "ideal" software development design to create the digital persona should: (i) foster engagement and satisfaction (ii) support multimodal approach by taking into account human feedback for an enhanced experience (iii) respect individual user behavior during interaction (iv) reflect the aim of creating the digital persona and (v) be ethically compliant to reflect trustworthiness and avoid misuse. Keeping these principles in mind, we present the immersive software development design approach implemented for the creation of the digital persona in our use case of creating the deepfake to enable art education (see Fig. 1).

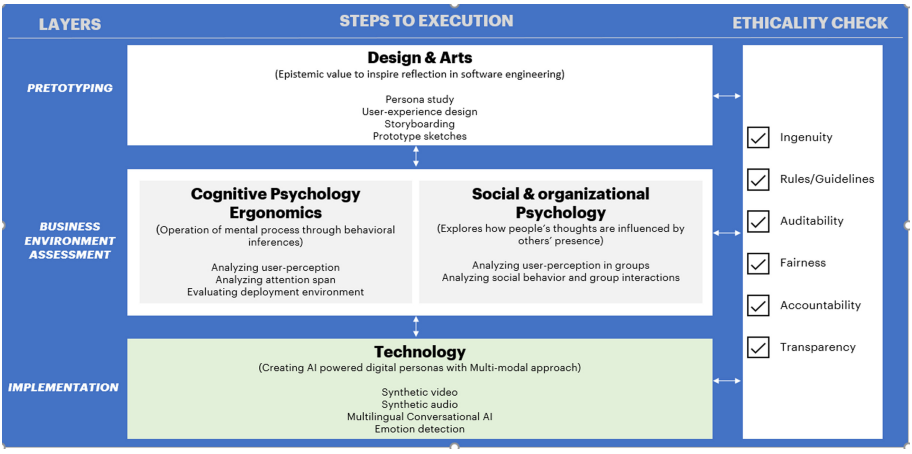

Fig. 1. Immersive software development design for creation of digital persona with multimodal inputs

3.1 Technology

Technology represents the software techniques used in generating the digital persona. We implemented 1. Face superimposition through deep learning to create the synthetic video 2. Speech synthesis using deep learning to create the synthetic audio and 3. Natural Language Understanding and Natural Language Processing

for Dialog Management 4. Real time Emotion analytics through facial analysis. All these components are provided as an input to the adaptation engine. The way these technology elements interact and intervene is shown in Fig. 2.

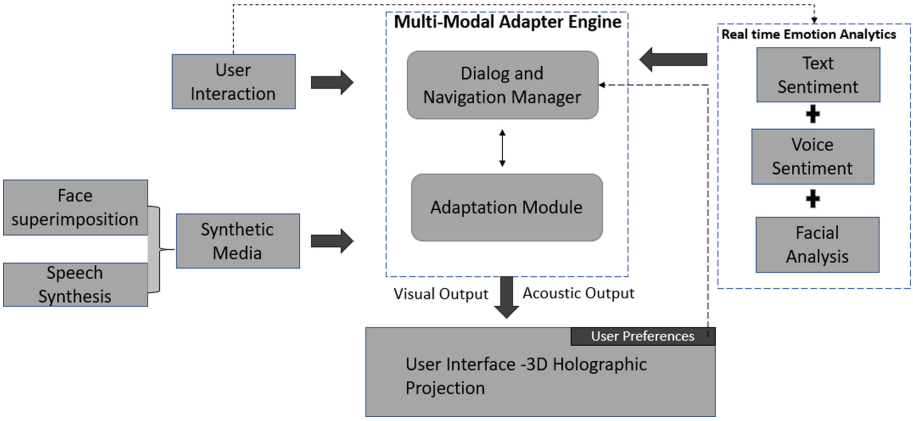

Fig. 2. Technology landscape depicting the interactions

Deepfakes are synthetic videos created using deep learning techniques where an individual's face is inserted into an existing video to appear as if the individual performed actions that occurred in the original video (Face Superimposition). In the generated video, the computer generated speech can also be integrated for a fake audio. The recent advances in speech synthesis leverage deep learning techniques to generate natural sounding speech.

In our implementation we created the synthetic video and audio and merged them together. The media file output was then marked with a digital signature to ensure appropriate usage. Several such deepfake media files (.mp4 format) were created based on the conversation topics.

To enable seamless user interaction with the persona during the conversation, we implemented the dialog and navigation manager to provide appropriate responses to user queries. The responses were in the form of the appropriate deepfake media files created. User preference on the topic of conversation was also provided to the dialog manager through the user interface. Real time emotion analytics was also implemented to analyze user reaction during the conversation and accordingly adapt the conversation to suit the user's mood. Emotions were aggregated based on the inputs of voice sentiment, text sentiment and facial emotions. For text sentiment, user's speech during the conversation was converted to text and then sentiment analysis was applied on the generated text.

Thus, based on inputs from the dialog manager, user preference and real time user emotions, the adaptation engine provided an appropriate deepfake video as an output which was projected to the user as a 3D Holographic projection [37].

Face Superimposition. For face superimposition, many freely available software's like deepfakes webβ [3], FaceSwap [15], FaceSwap-GAN [16] and DeepFaceLab [12] can be explored. For our design implementation, we picked up DeepFaceLab. For the implementation of deepfake in our work, we had to ensure the face superimposition works even if there is lot of movement.

The way in which all these software create fakes is through the trick of training two autoencoder networks, with shared weights of encoder network (i.e. a common encoder) and separate decoders for source and target face as shown in the Fig. 3. For the generation part, these architectures simply pass the latent representation of source face obtained from common encoder to the target face decoder as shown in Fig. 4, thereby converting the source face to target face in the process.

The faces that are passed as input to the common encoder for both training and generation are the aligned faces, that can be got through the use of any of the common face alignment algorithms like MTCNN, S3FD etc.

Other enhancement can also be added to make the output quality better. For example, jointly training a discriminator network along with the above encoder decoder network. Here the decoder network acts as generator and these are known as GAN based deepfakes. We can also try to increase the facial mask size to cover more of forehead for replacement.

In our case of creating the digital persona of the deceased artist, we had to superimpose the artist's face [Face A] on the face of an actor [Face B]. To ensure the face superimposition looks visually perfect, we realized based on our experiments that the actor needed to have the same facial structure, skin tone, hairstyle and beard.

The other guidelines for perfection include:

1. Sufficient Face A training data with a certain criterion: To collect the training data, various videos from YouTube featuring the artist were collected with the artist front facing the camera and without any obstructions on the face
2. In these video collections, the artist had to be of the same age and the lighting in the videos had to be consistent. Also, in the videos, the artist must be front facing the camera in most of the frames. With the mentioned criteria, we could collect around 10 min of video which lead to an extraction of around 9000 image frames.
3. Sufficient FaceB training data with a certain criterion: To get the training data, we had to video shoot the actor for a duration of around 10 min. During the shoot, we had to ensure the actor is both front facing and side facing the camera. The actor also had to exhibit varied expressions and the lighting during the shoot had to be only in the front.
4. Preprocessing of training data: Once the videos of the artist [Face A] from YouTube were collected, we had to remove the frames consisting of other people using a video cutter. The cut portions of the video were stitched back together before using it as input to the DeepFaceLab software.

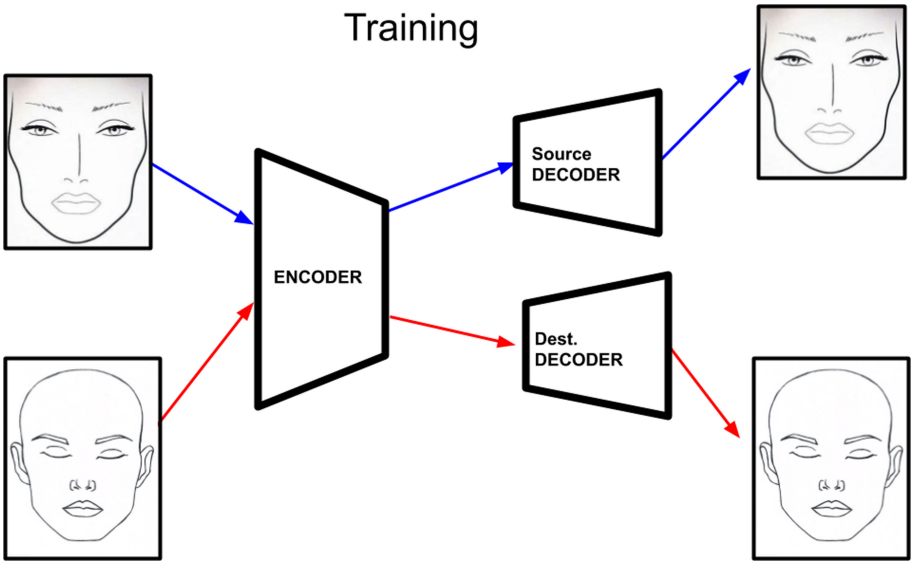

Fig. 3. Autoencoder architecture for deepfake training

Speech Synthesis. Since the person whose digital persona is to be created is deceased, we also had to synthesize his voice. To implement the voice synthesis, we explored the following software: a. sprocket-Kobyashi [26] b. ParallelWave-GAN [9] c. Tacotron2 [34] d. Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis [22] e. Neural Voice Cloning With Few Samples [8] and f. Deep Voice 3 [13,31].

The challenge we had with implementing speech synthesis was that we had limited samples of the artist's voice. Most of the software's listed above needed around 20 h of training data. With the challenge of low resource domain, we discovered that, with Deep Voice 3 we could retrain on target speaker data by training with 20 to 30 min of voice data from any other speaker. Therefore, from three prerecorded interview videos of the artist (available in YouTube), audio in .mp3 format was extracted and then 283 samples of audio files comprising of 5 s, 6 s, 9 s & 10 s were generated. In total, we could gather around 29 min of voice data that only contained artist's speech discarding other ambient noise. Along with the audio samples, corresponding text had to be extracted from each sample for training. We used Google's speech recognition API to extract text. However, the text output was not 100% due to the speaking style of the artist. Therefore, we had to manually curate the text data by listening to each audio sample. This was a time intensive task because audio samples had to be listened multiple times to record corresponding text accurately.

Post creating the speech synthesized audio, we merged the audio with the deepfake video created as described in the previous section. During the creation of the deepfake video, we ensured that the actor spoke as per the decided

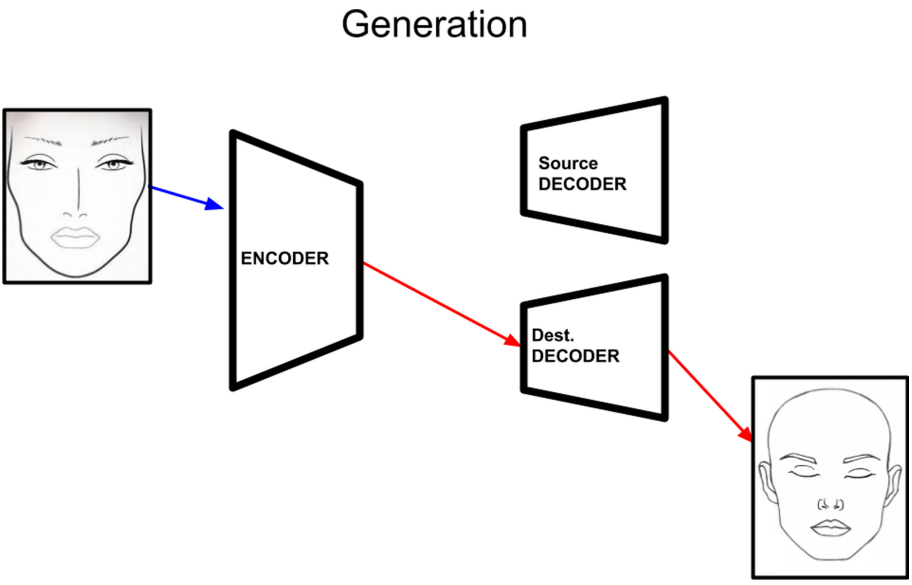

Fig. 4. Generating deepfakes after training

conversation topics. Thus, during merging of the deepfake video (synthesized video) and fake audio (synthesized audio), lip synchronization was taken care of.

Conversation AI. As mentioned earlier, the requirement of the digital persona was to have a conversation with the end user about Indian art and artworks. To enable this, any natural language processing or natural language understanding technologies [21] can be implemented to understand the user's intent and respond in the most useful way. We implemented Microsoft speech recognition service and Miscrosoft LUIS bot [4] API in conjunction with an adapter engine. The conversation was designed in a question-answer format. For the conversation with the persona, we configured 69 intents with 1244 utterances and Luis provided a 96.5% accuracy of predicting the intents correctly. In the engine, the intent response of Luis was mapped to a specific deepfake video. Thus, for any question asked by the user, a specific deepfake video of the digital person answering the question was provided as a response. The response video (output media file in mp4 format) was then projected on the 3D-holographic stage.

Emotion Analytics. Emotion Analytics software collects data and performs analysis on how people communicate verbally and non verbally to understand the mood or emotions of the person at an interval of time [27]. To avoid monotonous conversation between the digital persona and the end user, emotion analytics was implemented in the design approach. Various emotions of the user were detected as an aggregate of the outcome of voice sentiment, text sentiment and

facial emotions. Based on the involvement level and emotion detection, the aim was to change the conversation to suit the end user's mood. Involvement refers to the measure of user engagement during the conversation with the digital persona. For example, if the detected involvement level drops to less than 50%, the conversation must change to a different topic to lighten the mood or grab the attention of the user through a different dialog response.

We used microsoft azure's facial emotion detection and text sentiment analysis api's [2,7] for capturing facial and text emotions and used an open source implementation for capturing voice sentiments [5]. We then observed each short interaction with the digital persona and manually labelled it as high or low involvement. These api's collectively returns 18 real valued features, each between 0 and 1 representing strength of emotion. Text emotion analysis returns 3 features depicting positive, negative or neutral text emotions, facial emotion detects 8 features depicting anger, contempt, disgust, fear, happiness, neutral, sadness or surprise facial emotions and audio emotions returns 7 features depicting joy, surprise, neutral, anger, disgust, sadness or fear audio emotions. We collected around 700 such data points and used them to train a rbf kernel SVM [11]. We then use the trained SVM for predicting user's emotions. We take consent from users and constantly monitor them through a camera feed which is fed to three emotion detection APIs. The output from these api's are then passed to SVM classifier which then predicts high or low involvement along with prediction score.

Below is the snippet of sample data used to train the SVM classifier in Table 1.

3.2 Cognitive Psychology and Ergonomics

Cognitive psychology often uses computers to study human mental phenomena for the sake of developing general theories about human mental behavior. Cognitive ergonomics studies mental phenomena and applies theoretical knowledge in order to solve the practical problems related to using computers [1]. Cognitive ergonomics is based on psychological phenomena, such as knowledge, perception, and planning. It focuses on making systems simple and easy to use by analyzing human perception and mental processing.

By understanding cognitive psychology and ergonomics, we can build systems by (i) understanding user behavior and needs (ii) measuring user satisfaction (iii) planning content by: understanding how information is organized and presented to the users (iv) facilitating actions which are easy to access and use (v) focusing on accessibility as to how a disabled person can use the system (vi) collecting user feedback and improving the system functionality. To build the digital persona with the above stated pointers in mind and formalize a design approach, we had to understand the deployment environment - how will user interact with the persona? What will likely be of interest to them? And how the digital persona should adapt to support user's needs? To choose a design approach, we also had to think of the relevant criteria to assess and evaluate the approach for design

Table 1. Sample data used for training the SVM classifier

Text			Facial								Audio							Label
.82	0.01	0.17	0	0	0	0.05	0.62	0.28	0	0.05	0.7	0	0.23	0.07	0	0	0	High
0.72	0.02	0.26	0	0.05	0.06	0	0.19	0.6	0.02	0.08	0.6	0	0.32	0.08	0	0	0	High
0.2	0.7	0.1	0	0	0.1	0	.3	.5	0	.1	0	.1	.8	0	0	.06	.04	Low
.1	.2	.7	0	0	0	.1	.7	.2	0	0	.3	0	.5	.2	0	0	0	High
.3	.2	.5	0.7	0	0	.2	0	.1	0	0	0	0.1	.7	0	0.2	0	0	Low
.9	0	.1	0	0	0	0	.3	.6	0	0.1	.8	0	0	.1	0	0	0.1	High
0	.19	.81	0	0	0	0	0	.9	0.1	0	0	.9	.1	0	0	0	0	High
.1	.8	.1	0	0	0.3	0	0	.7	0	0	.1	0.4	.4	0	0	0.1	0	Low
.63	.3	.07	0	0.1	0	0	.6	.3	0	0	0	0.2	.6	0	0	0	0.2	High
0	.42	.58	0.3	0	0	0	0	.68	0.02	0	0	0	.8	.2	0	0	0	Low

representation and how well the approach answers different design questions by also allowing room for improvement.

Post our design thinking sessions and evaluating various storyboards, we chose to implement the digital deepfake person as a holographic projection [19]. There were 3 options for a holographic experience – a holographic table, holographic wall and a holographic stage. If a holographic table had to be used, the persona had to be projected only up to 3 ft. Because of this limitation, we did not consider this option. We further evaluated the options of a holographic wall and a holographic stage. The holographic stage setup gave us the flexibility of having a stage where we could place art artifacts on top of the stage for a more immersive effect. Also, the stage provided an auditorium environment which would enable large number of visitors experience the interaction with the digital persona of the deceased artist. The distance of the projection was designed to ensure people sitting on wheelchairs also could watch the experience without interruption. Since the distance was selected considering the accessibility, we had to choose to perform deepfake by not only taking video of the actor till the chest but the entire body. Implementing holographic projection of the entire length of the body was also visually more appealing than projecting it till the chest.

Conversation with the digital persona was designed to be related to the artist, Indian art and art works. The natural language understanding and natural language processing techniques were built in with multilingual feature to support both the global and local visitors in the museum. Accurate natural language understanding (NLU) is also a challenge due to speech recognition errors occurring due to noisy environment, speaker accent, speaker interruption, self correction etc. To minimize the errors, a moderator was needed to have a controlled environment and a noise reduction handheld mic with master/slave configuration was provided to the users so that there could be only one single user interacting with the persona at any given point in time. To facilitate user feedback and adapt to individual user's attention span, emotion analytics was implemented as described in Sect. 3.1.

3.3 Social and Organization Psychology

While cognitive psychology and ergonomics is about studying human behavior in a single user environment, social and organization psychology is concerned with the study of interpersonal, group, intergroup and interorganizational behavior. It deals with how interactions can be understood and modified with various types of interventions.

For applying social and organization psychology in creating the digital persona, we had to get views of the deployment environment from our internal team as well as the team in the museum. The museum team also needed to have the infrastructure visualized for the holographic projections. A pilot was also conducted with mock visitors to the museum and get their views on how they would feel about the interaction with the persona and what genre of questions would they likely ask the persona during their interaction to learn about the artist and Indian art. Group behavior which involved interactions among individuals post their experience in the pilot was also studied. In terms of organization psychology, the teams to build the digital persona was dynamic multi team environment. Thus, task design played an important key role in team effectiveness. The development environment was set to have openness and transparency among all key stakeholders and ensured no individual team worked in silos.

3.4 Design and Arts

Another traditional idea within human centered computing is formalizing the Arts as a source of inspiration. Certainly, artistic research can have an epistemic value and artworks can inspire reflection even in software engineering [35].

According to Solder Pold's book of Interface Aesthetics [32], he recognizes three types of works dealing with alternative realism of the interface which can be combined in the same artwork - a. Artworks dealing with Functional realism b. Artworks dealing with Media realism and c. Artworks dealing with Illusionistic realism.

In the creation of the holographic digital persona, artworks dealing with illusionistic realism is applied where interfaces are designed by maximizing reality towards immersive simulation. Applying this artwork design will enable users to become immersed in the illusionistic world presented. During storyboarding, sketches and visualization of the artist's projection on the holographic stage was made even before the implementation of the persona. This ensured early feedback and agreement of the outcome with the involved stakeholders. Sketches were made keeping in mind the considerations of the physical features, dress, style, props, shoe wear etc. Also, to maximize the immersive simulation in the outcome, post production tools to provide the digital visual effects were used.

3.5 Ethics

Since the intention is to apply deepfake for good, it must be created ethically as well. The challenge is how do we call a fake as genuine? If we follow the rules of

fairness, accountability and transparency, we can address some of the concerns related to the fakes. Accountability and transparency mean to report, explain the algorithmic decision making, secure training data, as well as mitigate any negative social impact. To minimize negative social impact, secure communication channels need to be used and the deepfakes need to be deployed in a controlled environment with authorized access to play the deepfake video. Transparency can be achieved by making the data and code available along with the decisions taken during the system development. For example, the decisions taken for choosing a particular software for face superimposition and speech synthesis, decision taken on the parameter values for the face superimposition and speech synthesis etc.

Fingerprinting can be used with encryption while creating the video in order to ensure authorized use during deployment. Permissions and consent need to be taken from source artist or from the rightful heir in the artist's family of whose fake we want to create. To ensure fairness, we must evaluate that the person chosen to be source of the fake does not provide an impression of gender and racial bias when the outcome of the deepfake video is consumed by the end users.

Ingenuity in AI and technology development must be tempered with integrity, values, governance, policy and legal consequences for misconduct. These are issues that need collectively informed public solutions with input from industry, consumer and government regulators [33]. The questions we need to answer are: a. Have we taken consent from both the parties (actor and the artist) b. Does the message put forth though the deepfake is trustworthy and is reliable source of information? c. Does the actor and the message put forth have no indications of any gender or racial bias? d. Can we confidently showcase the deepfake created in public or is it only for private viewing? e. Do we have enough reasoning to explain the creation of the deepfake f. Can the deepfake creation process be auditable? g. Are there any existing guidelines or regulatory compliance followed for creating deepfake for good?

The deepfake should be ethical by design itself. An industry wide commitment to basic legal standards, significant regulation and technological ethics need to be implemented to address immediate harms of any bad design. For example, from regulatory point of view, the content distribution platforms should be put under certain stricter legal obligations to control the deepfake content dissemination.

4 Outcome

We applied our digital persona generation approach to generate the 3 dimensional holographic persona of M.F. Husain (known as Picasso of India) with an aim to generate interest in art education for children.

We implemented two versions of the same. One, the online version that allows users to interact with persona through a web browser. Here is the link of a short demo video for the online version.

The other is the visually appealing 3 dimensional holographic version of M.F. Husain. Here is the link of how the experience would look when the video is

projected on the 3d holographic stage. When viewed by naked eye, the live experience is immersive. However, the 3 dimensional aspects do not come out well in the video link presented above since it is shot using a camera and whole experience looks 2 dimensional as the depth can't be realized.

In both the versions, the persona converses with users on his life and artworks to invoke interest in the field of art. It also speaks in the voice of M.F. Husain. The experience is as if the deceased artist himself speaks to you.

The key outcomes of the generated persona are:

1. Visual appeal - We and the team at the museum found interacting with the 3 dimensional holographic digital persona more engaging as compared to interaction using 2d screen [29]. Also, since the persona converses in the voice of deceased artist, the experience is as if artist himself came to life.
2. Conversational - The persona answers around 1244 variations of questions asked by visitors related to his life and artwork. For example, questions related to persona's favourite painting, painting style, childhood, entry in the field of art, art journey etc. It can also answer tactfully to diplomatic questions to keep conversation light and funny.

5 Discussion

We have presented so far, the human centered computing design approach implemented for the creation of the digital persona. For our business implementation scenario of creating the digital persona of an artist to enable art education, the system was designed as multimodal with inputs and outputs in more than one modality. It was designed to be proactive where the digital persona understood the cultural and social context to provide an appropriate response. Affective computing was implemented to recognize, interpret and process the human interactions.

Clients or businesses who wish to use generative AI techniques to create digital personas for good, face gap in lack of reference framework, agility, adaptability and compliance issues. They need an understanding of digital persona creation to bring in the best experience for the end user. We believe, adapting the human centered design principals described in the above sections will help business to rapidly develop digital persona while being compliance aligned to provide material business impact and create sustainable value.

In future, we plan to improve the feedback loop by improvising on the text sentiment analysis and voice tone analysis on the conversation between the user and the persona.

6 Experiment Results

To evaluate the impact of our framework on learning experience of kids, we conducted a small social experiment post the deployment of the digital persona at the museum. The social experiment was conducted between two groups of

10 kids each, between the age group of 9–14 years. We wanted to check if this experience of interacting with a 3d Holographic digital persona would bring in improvement in retention of memory and enhance creativity among kids. Not only the experiment results showed a positive impact on both the factors, but also the kids who were exposed to this experience were motivated to visit the museum again.

The details of the social experiment is as follows: Both the groups were taken to museum and were taught about legendary Indian artist M.F Husain. Both the groups were given placards of 20 questions that they can ask and were encouraged to ask about Husain's life, family, artwork, interest, favorite paintings. The kids were also told that there could be a test from these 20 questions after few days (so that they can ask more questions). First group learned about M.F Husain through the traditional way in which a museum guide walked them through the museum while telling about Husain's paintings, life and interests. Kids were also allowed and encouraged to ask questions. Second group interacted with digital persona of M.F Husain and was also accompanied by a museum guide. We observed that the second group retained the learning better than first group, they also rated their experience higher than the first group. The total number of questions asked by first and second groups respectively was 14 and 17. Each group was asked to rate their experience on a scale of 1 to 4, i.e. how satisfied they were on learning about Husain from four choices 4) Extremely Satisfied 3) Satisfied 2) Dissatisfied 1) Extremely Dissatisfied. First group average score was 3 and second group average was 3.67.

After 3 days, all students were given a test of 14 short questions. All of which were parts of the elaborate answers provided to them for 20 questions. First group answered 34% correctly, while second group answered 43% correctly. Each kid was also asked the same questions that they had asked in their previous visit. While first group correctly remembered 6 answers out of 14 questions asked i.e. 43%, second group remembered 10 answers out of 17 questions asked i.e. 59%. All the kids were also asked if they would like to have the experience again, only six from first group answered in a yes, while eight kids from second group wanted to interact with Husain's persona again.

7 Conclusion

For a richer and immersive experience, any software system should employ the human centered computing design approach. Human centered computing involves creation of theoretical frameworks, design and implementation of technical approaches and systems in many areas which include – a. Systems for problem-solving by people interacting in distributed environments b. Multimedia and multimodal interfaces c. Intelligent interfaces and user modelling d. Information visualization and adaptation of content to accommodate different capabilities, modalities, bandwidth, latency etc. [10].

In this paper, we showed the human centered computing design approach implemented for the creation of a digital persona adhering to the characteristics

of being expressive, responsive, proactive and adaptive. While implementing the approach, it is observed that sociology, psychology and cognitive science play important roles apart from just the use of technology in presenting the digital persona for an engaged and immersive experience. We provided the outcome of our approach for creating digital persona of a renowned deceased artist M.F. Husain. We also conducted and presented the results of a social experiment which showed that our implementation allows children visiting the museum to explore artists and art forms in completely new ways. This framework allows us to create human centric digital personas in an affordable manner. These digital personas can be used in various other industries as well to improve customer experience.

References

1. Cognitive ergonomics and user interface design (2008). http://members.upc.nl/g.haan24/articles/chapter1.html. Accessed 7 Feb 2020
2. An AI service that analyses faces in images (2020). https://azure.microsoft.com/en-in/services/cognitive-services/face/. Accessed 17 May 2020
3. Create your own deepfakes online (2020). https://deepfakesweb.com/. Accessed 5 Feb 2020
4. Language understanding (LUIS) (2020). https://www.luis.ai/home. Accessed 15 June 2020
5. Real-time multimodal emotion recognition (2020). https://github.com/maelfabien/Multimodal-Emotion-Recognition. Accessed 29 July 2020
6. The Salvador Dali museum (2020). https://thedali.org/. Accessed 17 July 2020
7. Text analytics API documentation (2020). https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/. Accessed 22 July 2020
8. Arik, S.Ö., Chen, J., Peng, K., Ping, W., Zhou, Y.: Neural voice cloning with a few samples. CoRR abs/1802.06006 (2018). http://arxiv.org/abs/1802.06006
9. kan bayashi: Unofficial parallel wavegan (+ MelGAN) implementation with Pytorch (2020). https://github.com/kan-bayashi/ParallelWaveGAN. Accessed 24 Feb 2020
10. CISE - IIS - About (2020). http://www.nsf.gov/cise/iis/about.jsp. Accessed 12 Jan 2020
11. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
12. DeepFaceLab is the leading software for creating deepfakes (2020). https://github.com/iperov/DeepFaceLab. Accessed 11 Jan 2020
13. Pytorch implementation of convolutional neural networks-based text-to-speech synthesis models (2019). https://github.com/r9y9/deepvoice3_pytorch. Accessed 21 Feb 2020
14. Egger, M., Ley, M., Hanke, S.: Emotion recognition from physiological signal analysis: a review. Electron. Notes Theor. Comput. Sci. **343**, 35–55 (2019). https://doi.org/10.1016/j.entcs.2019.04.009. http://www.sciencedirect.com/science/article/pii/S157106611930009X. The Proceedings of AmI, The 2018 European Conference on Ambient Intelligence (2018)
15. Faceswap: Deepfakes software for all (2020). https://github.com/deepfakes/faceswap. Accessed 29 Jan 2020

16. A denoising autoencoder, adversarial losses and attention mechanisms for face swapping (2019). https://github.com/shaoanlu/faceswap-GAN. Accessed 19 Jan 2020
17. Fried, O., et al.: Text-based editing of talking-head video. ACM Trans. Graph. **38**(4), July 2019. https://doi.org/10.1145/3306346.3323028
18. You can now speak using someone else's voice with deep learning, July 2019. https://towardsdatascience.com/you-can-now-speak-using-someone-elses-voice-with-deep-learning-8be24368fa2b. Accessed 16 Feb 2020
19. Holography (2020). https://en.wikipedia.org/wiki/Holography. Accessed 4 Feb 2020
20. Jaimes, A., Sebe, N., Gatica-Perez, D.: Human-centered computing: a multimedia perspective. In: Proceedings of the 14th ACM International Conference on Multimedia, MM 2006, New York, NY, USA, pp. 855–864. Association for Computing Machinery (2006). https://doi.org/10.1145/1180639.1180829
21. These five platforms will make your bots language-intelligent (2016). https://chatbotsmagazine.com/these-five-platforms-will-makeyour-bots-language-intelligent-634556750abd. Accessed 5 Jan 2020
22. Jia, Y., et al.: Transfer learning from speaker verification to multispeaker text-to-speech synthesis. CoRR abs/1806.04558 (2018). http://arxiv.org/abs/1806.04558
23. Kazeminia, S., et al.: GANs for medical image analysis. Artif. Intell. Med. **109** (2020). https://doi.org/10.1016/j.artmed.2020.101938. http://www.sciencedirect.com/science/article/pii/S0933365719311510
24. VGGFace implementation with Keras framework (2020). https://github.com/rcmalli/keras-vggface. Accessed 4 Jan 2020
25. Kietzmann, J., Lee, L.W., McCarthy, I.P., Kietzmann, T.C.: DeepFakes: trick or treat? Bus. Horizons **63**(2), 135–146 (2020). https://doi.org/10.1016/j.bushor.2019.11.006. http://www.sciencedirect.com/science/article/pii/S0007681319301600
26. Kobayashi, K., Toda, T.: sprocket: Open-source voice conversion software, pp. 203–210, June 2018. https://doi.org/10.29007/s4t1
27. Emotion analytics (2018). https://searchcustomerexperience.techtarget.com/definition/emotions-analytics-EA. Accessed 14 Jan 2020
28. Nguyen, T., Nguyen, C., Nguyen, T., Nguyen, D., Nahavandi, S.: Deep learning for deepfakes creation and detection, September 2019
29. 10x your employee engagement with immersive learning experiences (2019). https://www.jolt.io/blog/10x-your-employee-engagement-with-great-learning-experiences. Accessed 27 June 2020
30. van den Oord, A., et al.: WaveNet: a generative model for raw audio. arXiv (2016). https://arxiv.org/abs/1609.03499
31. Ping, W., et al.: Deep voice 3: scaling text-to-speech with convolutional sequence learning (2017)
32. Pold, S.: Interface realisms: the interface as aesthetic form. Postmod. Cult. **15**, January 2005. https://doi.org/10.1353/pmc.2005.0013
33. Sanders, N., Wood, J.: The Humachine: Humankind, Machines, and the Future of Enterprise. Taylor & Francis, Abingdon (2019). https://books.google.co.in/books?id=OVauDwAAQBAJ
34. Shen, J., et al.: Natural TTS synthesis by conditioning WaveNet on MEL spectrogram predictions, pp. 4779–4783, April 2018. https://doi.org/10.1109/ICASSP.2018.8461368

35. Tomáis, E.: How the arts can help tangible interaction design: a critical re-orientation. Informatics **4**, 31 (2017). https://doi.org/10.3390/informatics4030031
36. Wang, Y., et al.: Tacotron: a fully end-to-end text-to-speech synthesis model, March 2017
37. Liang, W.: The 3D holographic projection technology based on three-dimensional computer graphics, pp. 403–406, July 2012. https://doi.org/10.1109/ICALIP.2012.6376651