Real or Rendered: A Comparative Study of Fidelity in AI-Generated Avatars

Lamija Halilović
Supervised by: Selma Rizvić, Aya Ali Al Zayat

Abstract

With the development of generative artificial intelligence, the distinction between AI-generated content and human reality is becoming increasingly blurred. This research explores the current state of easily accessible generative AI tools by evaluating the visual and vocal authenticity of digital humans. The study features a comparative analysis of eight AI video generation tools, each used to produce an identical video sequence. These videos, along with one real human video, were presented to a diverse group of participants to examine their ability to distinguish between synthetic and real media. Key factors such as facial expressions, lip-sync accuracy, and the presence of the uncanny valley effect were examined throughout the study. The evaluation indicates that even users familiar with artificial intelligence were misled by the high fidelity of avatars, thinking that they were real human beings. Certain AI tools have overcome the uncanny valley, making high-fidelity digital humans natural to the observer. The results mentioned above suggest that the realism of synthetic media has bypassed human sensory perception, highlighting the urgent need for advanced AI verification tools.

Keywords: Generative AI, Visual Fidelity, Uncanny Valley, Human Perception

1 Introduction

Generative artificial intelligence is evolving rapidly, to the point where it is becoming difficult for everyday users to distinguish between synthetic and real content. AI-generated avatars are increasingly being integrated into education, government, corporate environments, and social networks [19][18]. These digital humans not only look like real people but also sound and act like them.

Although high-end computing power and expensive software can produce nearly indistinguishable results[13][15], this study focuses on tools accessible to the average user with standard hardware and software. The AI applications analyzed in this study are either free or offer a free trial, with the intention of exploring the creative capabilities available to the general public. Common applications with which people interact daily, such as Google Gemini used in this study, have become so advanced that they are even now capable of generating highly realistic AI avatars [6]. On the other hand, there are many AI tools built for photo and video generation or specifically for AI avatar creation. The mentioned applications use photos and voice recordings as sources, creating digital characters identical to humans. The avatars produced can be placed in different environments and their appearance and voice can be modified, giving creative freedom to users.

Numerous AI tools use the same underlying foundation, such as Kling AI [12]. By integrating Kling's API, these platforms customize the generation process to achieve different results. Users can use a similar methodology to develop avatars on their own devices, which can give incredible results in combination with quality hardware. However, this study focuses on existing tools; thus, this method will not be tested. The API adaptation methodology leads to varying levels of quality and realism in different tools. This study focuses on comparing various generation AI tools, specifically evaluating the visual and vocal fidelity of the avatars.

In the creation of highly realistic digital humans, there is a tendency for them to become realistic to the point where they evoke a feeling of discomfort in viewers. This phenomenon is commonly known as the uncanny valley, introduced by Mori et al. in 1970 [14]. In terms of artificially generated avatars, the uncanny valley would be triggered if a realistic avatar has a robotic voice. Along with the obvious triggers, there are micro-behaviors, for example, blinking or other facial twitches. Although often not perceived by the observer, these involuntary behaviors are a sign to the brain to verify the human presence. The positional standing of generated avatar along the uncanny valley curve is highly dependent on these micro-behaviors[14].

Given the rapid pace of development in this field, the objective of this study is to provide information on the current capabilities of publicly accessible AI tools, evaluated using visual and vocal fidelity criteria. As the abilities of these tools expand, such periodic estimations remain valuable for understanding what is realistically achievable by the average user.

Under the impact of increasing synthetic media realism, this study also addresses the ethical implications of AI avatar generation, including the risks of identity misappropriation and digital deception.

This study offers three key contributions: a comparative analysis of eight freely accessible AI avatar tools tested under controlled and identical conditions, a perceptual evaluation framework examining both visual and vo-

cal fidelity dimensions, and preliminary empirical insight into whether current consumer-grade tools have moved beyond the uncanny valley threshold.

The following sections provide an analytical review of existing research, followed by a case study outlining the generation methodology. Subsequently, the results of the user evaluation are presented. The paper concludes with a discussion of ethical issues of AI avatars and a summary of key findings and future work.

2 Related work

In terms of theoretical analysis, the previously mentioned work of Mori et al. written in 1970 had a great impact on understanding human computer interaction. They described the relationship between the human likeness of an entity and the perceiver's affinity for it. As human likeness increases, the affinity generally increases. In this phase, industrial robot and stuffed toy are used as an example. Then it drops sharply into the valley with near-human resemblance, a zombie representing the lowest point of the valley. It rises again at full human likeness, reaching the highest point that represents a healthy human. There are two cases of mentioned graph, one which takes movements into consideration and the other does not. Taking motion into account increases the valley. Using a prosthetic hand as an example, they explain that when seeing something that represents part of a living being, there is the expectation that it also feels like a real human being. The coldness and firmness of the prosthetic hand are what triggers the feeling of uncanniness in observers, placing it in the valley on the graph [14]. This principle can be applied to generated avatars, where observers expect that a realistic avatar talks and acts as a real human being. The authors' analysis contributed to a better understanding of the user evaluation and possible causes of discomfort in observers.

A recent study by Baake et al. indicates that visual realism significantly improves the perceived trustworthiness of AI-generated personas compared to more abstract or stylized designs in science communication. In addition to realism, the gender of the avatar had an impact on the results. Male avatars were perceived as more competent than female avatars, though this did not apply to other dimensions of trustworthiness. Although the results indicate that discomfort was not triggered by highly realistic avatars, the authors concluded that the reason for this could be that the realism was not at a high enough level to evoke uncanniness [3]. This research indicates that newer and more advanced AI technologies could potentially surpass the uncanny effect. However, trust in science was a circumstance that had a great impact on observer's likeness of the created avatars. In non-formal environments, there is still the possibility of the presence of the uncanny valley effect. These findings had an impact on the visual style of the AI avatar used in this study, emphasizing the importance of balancing realism and acceptability to achieve optimal perceived trustworthiness.

Another recent analysis focuses on three user perceptual interaction experience-related indicators: attractiveness, trustworthiness, and eeriness. A total of 13 studies were included in the timeline from 2015 to 2025 and 2343 participants. In terms of attractiveness, high realism avatars received significantly higher ratings than medium and low counterparts. Similarly, trustworthiness levels in the higher realism avatars were greater than in the lower realism avatars, while medium realism showed no significant difference in comparison to the other. The eeriness ratings were highest received by the medium realism avatars, which is consistent with theoretical predictions, given that partially human-like features tend to disrupt the observer's perceptual expectations [16]. These findings further supported the options based on their visuals, as they highlight that high realism avatars not only enhance perceived attractiveness and trustworthiness, but also minimize eeriness.

The use of AI avatars in postoperative patient education was investigated by Haider et al. In the study, patients were chosen based on their prior clinical relationship with the surgeon. This approach was chosen to assess the specific value of personal familiarity rather than the perception of a generic medical avatar. The preliminary findings of this study suggest that clinical avatars do not necessarily require perfect realism to avoid triggering eeriness with an established patient-physician relationship. The avatar system relied on pre-recorded responses rather than real-time AI generation, a deliberate methodological choice that prioritized safety and clinical accuracy by eliminating the risk of hallucinations [8]. This research indicates that recognizable avatars may avoid the uncanny valley if they sound and look like a real person who represents them. The findings of this study are particularly relevant to the present work, as they suggest that familiarity and perceived authenticity may play a more significant role in avatar acceptance than technical realism alone. This contributed to the consideration of how audience familiarity with the communicator could influence the effectiveness of the AI avatar used in this research.

3 Case study

This comparative study is based on the user evaluation of eight AI generated videos. The first step was conducting a research on the available tools for video generation. In the paper AI Avatars in Action: A Review of Applications, Challenges, and Future Potential, the authors listed the tools available for photorealistic avatars and digital twins [19].

The first tool found in this research was Vidnoz [17]. Vidnoz is a free generation tool with a daily limit of one minute per video. It features a simple interface where users can upload a photo and video description. Users

maintain the flexibility to choose between numerous versions, with Wan2.6 serving as the engine for this tool, whereas other base tools, such as Kling and Veo were used through other tools. This tool does not offer the generation of custom voice for the avatar.

AI Studio by Deepbrain is the second tool mentioned in the aforementioned paper[4]. This application does not have a free plan, but it offers a free trial. The process of creating an avatar is user-friendly. Users upload a photo and read a given script for voice generation. After avatar generation, the user can use the same avatar in multiple projects with the only adjustment being the script.

HeyGen was primarily found in the paper, but it was also suggested by numerous users on online forums [9]. This tool received the best reviews as it was found to be practical and efficient. HeyGen is a partially free tool, with three minutes of videos per month and limited features. The avatar creation process is simple where users input a photo and choose a voice from HeyGen's voice library. After creating the avatar, users can also customize it using AI, changing its appearance. The avatar can be used in multiple projects with the ability to change voice if desired. An additional feature not found in other tools was the option of adding a voice note to guide the avatar on how to deliver the script, giving the possibility to make the avatar sound more natural with tone variability.

Other tools used in this study were found by reading online forums and via browser and social media search.

Google Gemini is a well-known multimodal large language model (LLM) [6]. Although generally used for different tasks, Gemini offers an option of video generating using the Veo3.1 model. This feature is available to every user with a limit of three videos per day. The process is simple: the user uploads a photo and writes a prompt of what the avatar's action and speech should be. The voice of the avatar cannot be chosen explicitly, therefore, the results are unexpected.

Another Google tool is Flow [7]. Despite the fact that Gemini and Flow are both produced by Google and share a similar generating process, these tools were shown to give different results. Flow's interface is adapted to film making, where users can choose starting and ending frames, or add multiple elements as the ingredients for a video. Multiple videos can then be combined into a movie. Voice selection is also not an available feature in this tool, so the results remain unpredictable in that aspect.

Akool is a tool recommended by users as a good HeyGen alternative [1]. This application does not offer a free plan, but it offers a free trial. The interface is similar to HeyGen and is simple to use with an option to generate a voice based on the user's voice recording. The avatar created once can be used multiple times, in different prompts. This tool uses different API sources, one of them being Kling [12].

Qwen is a tool found in the user recommendations on the online forums [2]. Qwen is Alibaba's multimodal LLM, and the process of avatar generation is identical to Google Gemini.

FalcoCut offers an interface similar to HeyGen and Akool. The user uploads a photo and then selects a voice from different bases. FalcoCut is also an API based tool, using Kling [12].

The main principle of this study was to evaluate each tool under identical conditions and to generate eight identical videos for users to review. Essentially, every application mentioned above requires a photo upload. This photo, shown in Figure 1, was extracted from the source video, filmed by a real person, which was used in user evaluations to assess the users' ability to distinguish authentic human content.

Figure 1: Source Image Used for AI Avatar Generation

Voice generation is not a feature offered by every tool, only three of them, and the process of generation required a different approach in all three tools. Each application was tested using the same one-sentence script, shown in Table 1, ensuring that all time limits were taken into account. Some tools are prompt based, where users need to input instructions. Each AI based on prompts received the same set of instructions shown in Table 1, guiding it to deliver the script naturally and maintain the same avatar look as in the photo.

Script	**Prompt**
Hi, I'm Lamiya and today I'm testing several AI avatars to compare their facial expressions, lip movements and voices.	Delivery should be in a woman's voice, natural and with the right emphasizing of words (appropriate to context). If appropriate to context, she can also use her hands in explaining. The subject's appearance and scenery should remain the same.

Table 1: Script and Prompt Text Used For Video Generation

After preparing the source photo, voice notes, script, and prompt, the corresponding data was provided to all AI tools. The approximate time to generate a video does not include the preparation of the data but only the time it takes to get the finished video. Time is expressed in min-

utes, with ± 1 minute representing a possible inaccuracy due to internet connection or other similar factors. The general overview of the tools used is presented in Table 2.

AI Tool	Version	Source	Approximate time (minutes)
Gemini[6]	Veo 3.1	image	2
Flow[7]	Veo 3.1	image	2
HeyGen[9]	Avatar IV	image, audio	3
Akool[1]	Kling 3.0	image, audio	4
Qwen[2]	Qwen3-Max	image	2
FalcoCut[5]	Kling 2.6 Pro	image	4
Vidnoz[17]	Wan 2.6	image	2
AI Studios[4]	Kling 2.6 Pro	image, audio	4

Table 2: Overview of Generative Platforms Evaluated

The complete set of generated materials was assigned neutral labels differentiated solely by test case number, ensuring that the labeling did not influence the participants' evaluations.

4 User evaluation of avatars fidelity

4.1 Overall structure

A pilot study (n=7) was conducted in which participants viewed generated video materials and an original video that was filmed, and subsequently completed a structured questionnaire, rating their responses on a five-point Likert scale. In addition to rating each video, users had the opportunity to express which of the videos they think is real. The person in the video materials was not previously known by the majority of respondents (n=6). The participants evaluated each avatar according to eight perceptual criteria:

- naturalness of eye behavior (blinking, focus),
- realism of facial expressions (emotional precision),
- lip synchronization,
- naturalness of head and body movement,
- audio-visual alignment,
- speech tone and tone variability,
- naturalness of speech tempo,
- and overall perceived naturalness.

4.2 Demographic data

Of the seven participants, five identified as female and two as male. The majority of the participants (57.1%, n=4) belong to the age range 25-34, and the remaining participants are distributed across the 18-24, 35-44 and 55+ age groups, each represented by one participant. Six of the seven users are familiar with artificial intelligence and five out of seven indicated prior awareness of AI avatars.

4.3 Case 1: AI Studios

The first video evaluated by users was generated using AI Studios. The results of the evaluation are presented in Table 3.

Criterion	M(1-5)	SD
Naturalness of eye behavior	3	1.63
Realism of facial expressions	2.29	1.11
Lip synchronization	2.86	1.07
Naturalness of head and body movement	2.86	1.57
Audio-visual alignment	3	1.15
Speech tone and tone variability	2.29	1.11
Naturalness of speech tempo	2.43	1.40
Overall perceived naturalness	2.86	1.07

Table 3: AI Studios Avatar Evaluation

Six of the seven participants reported that the video evoked a sense of eeriness. A participant who stated his familiarity with artificial intelligence and AI avatars, belonging to the age group 35-44, believed that this video was a real person. One of the participants declared that they found the eyes and the intonation strange.

4.4 Case 2: Akool

The second video evaluated by users was generated using Akool. The results of the evaluation are presented in Table 4.

Criterion	M(1-5)	SD
Naturalness of eye behavior	3.14	1.46
Realism of facial expressions	3.43	1.40
Lip synchronization	3.57	1.51
Naturalness of head and body movement	3.71	1.50
Audio-visual alignment	4.14	1.21
Speech tone and tone variability	3.43	1.72
Naturalness of speech tempo	3.43	1.40
Overall perceived naturalness	3.43	1.51

Table 4: Akool Avatar Evaluation

The majority of the respondents found this video to be real (n=4), two respondents reported a feeling of eeriness. This sensation of eeriness was reported by two distinct profiles: a participant with a technical understanding of AI, and another from the 55+ age group who lacked familiarity with AI technology. One person was indecisive on

behalf of some elements appearing to be real, while others were not.

4.5 Case 3: Original video

The third video evaluated by users was the original video, used in the study to investigate whether users can distinguish a real person from the generated avatars. The users had the ability to evaluate the original video using the same criteria as the rest. The results of the evaluation are presented in Table 5.

Criterion	M(1-5)	SD
Naturalness of eye behavior	4.29	0.95
Realism of facial expressions	4.43	0.98
Lip synchronization	4.43	0.98
Naturalness of head and body movement	4.29	1.25
Audio-visual alignment	4.57	0.79
Speech tone and tone variability	4.43	0.98
Naturalness of speech tempo	4.14	1.07
Overall perceived naturalness	4.29	0.95

Table 5: Original Video Evaluation

This video was evaluated as real by all users.

4.6 Case 4: HeyGen

The fourth video evaluated by users was generated using HeyGen. The results of the evaluation are presented in Table 6.

Criterion	M(1-5)	SD
Naturalness of eye behavior	2.71	1.25
Realism of facial expressions	2.43	1.27
Lip synchronization	2.71	1.25
Naturalness of head and body movement	2.43	1.40
Audio-visual alignment	3	1.15
Speech tone and tone variability	2.57	0.98
Naturalness of speech tempo	2.43	0.98
Overall perceived naturalness	2.71	1.25

Table 6: HeyGen Avatar Evaluation

Five users reported this video as unnatural and two rated it as real. One participant who stated that the video is real is previously familiar with AI avatars, while the other is familiar with artificial intelligence, but not with AI avatars.

4.7 Case 5: FalcoCut

The fifth video evaluated by users was generated using FalcoCut. The results of the evaluation are presented in Table 7.

Criterion	M(1-5)	SD
Naturalness of eye behavior	3	1.41
Realism of facial expressions	2.57	1.51
Lip synchronization	3	1.53
Naturalness of head and body movement	2.43	1.27
Audio-visual alignment	3.29	1.50
Speech tone and tone variability	2.29	1.11
Naturalness of speech tempo	2.14	1.07
Overall perceived naturalness	2.71	0.95

Table 7: FalcoCut Avatar Evaluation

The majority of users evaluated this video as unnatural, while one user believed it was real. This user has stated their previous familiarity with artificial intelligence and AI avatars.

4.8 Case 6: Flow

The sixth video evaluated by users was generated using Flow by Google. The results of the evaluation are presented in Table 8.

Criterion	M(1-5)	SD
Naturalness of eye behavior	3.86	1.35
Realism of facial expressions	4	1.41
Lip synchronization	4	1.53
Naturalness of head and body movement	3.71	1.50
Audio-visual alignment	4.14	1.46
Speech tone and tone variability	3.14	1.57
Naturalness of speech tempo	3.43	1.62
Overall perceived naturalness	4	1.41

Table 8: Flow Avatar Evaluation

Four of the users found the video to be realistic and two reported it as unnatural. One of the users who reported it as unnatural is familiar with AI avatars, while the other person is not. One user stated that while the avatar does look real, its behavior seems overly rehearsed.

4.9 Case 7: Vidnoz

The seventh video evaluated by users was generated using Vidnoz. The results of the evaluation are presented in Table 9.

Criterion	M(1-5)	SD
Naturalness of eye behavior	1.86	1.21
Realism of facial expressions	1.43	0.79
Lip synchronization	1.43	0.79
Naturalness of head and body movement	1.57	0.98
Audio-visual alignment	1.57	1.13
Speech tone and tone variability	1.29	0.49
Naturalness of speech tempo	1.57	1.13
Overall perceived naturalness	1.43	0.79

Table 9: Vidnoz Avatar Evaluation

Every user reported that this avatar evoked the feeling of uncanniness.

4.10 Case 8: Gemini

The eighth video evaluated by users was generated using Google Gemini. The results of the evaluation are presented in Table 10.

Criterion	M(1-5)	SD
Naturalness of eye behavior	3.14	0.90
Realism of facial expressions	2.43	0.53
Lip synchronization	3.57	1.27
Naturalness of head and body movement	2.43	1.27
Audio-visual alignment	3.71	1.25
Speech tone and tone variability	2.86	1.35
Naturalness of speech tempo	2.57	1.40
Overall perceived naturalness	2.86	0.69

Table 10: Gemini Avatar Evaluation

Five users evaluated this video as unnatural, one user classified it as real. This user has previously stated to be familiar with artificial intelligence, but not with AI avatars. One participant stated that the lip synchronization in this video was unnatural.

4.11 Case 9: Qwen

The ninth video evaluated by users was generated using Qwen. The results of the evaluation are presented in Table 11.

Criterion	M(1-5)	SD
Naturalness of eye behavior	2.43	1.13
Realism of facial expressions	2.14	1.21
Lip synchronization	1.29	0.76
Naturalness of head and body movement	2.14	1.21
Audio-visual alignment	1.43	0.79
Speech tone and tone variability	1.57	1.13
Naturalness of speech tempo	1.71	1.25
Overall perceived naturalness	1.57	0.79

Table 11: Qwen Avatar Evaluation

The majority of the users (n=6) found this video to be unnatural, while one user stated that it looks realistic, but the quality of the camera is poor. This user is familiar with AI avatars.

4.12 Overall impression

In the concluding question, participants were asked to identify which video depicted a real human being. Five participants correctly identified avatar three as the real person. One person selected avatar four, generated using the HeyGen platform, as the most convincing. This participant, who belongs to the age group 18-24, stated that they are familiar with artificial intelligence, but not with AI avatars. Another person chose avatar six, generated using Flow by Google. This participant, who belongs to the age group 55+, stated no previous knowledge of artificial intelligence or AI avatars. One participant noted difficulty choosing between avatars two (generated using Akool) and three (the real person), and another stated that they would have selected avatar five (generated using FalcoCut), had they not been previously familiar with the real person's voice.

5 Ethical issues of AI avatars creation

Generated AI avatars are developing to a point where it is challenging to recognize whether the content is synthetic or not. This advancement in visual fidelity inevitably brings forth ethical concerns. The most obvious is unauthorized identity synthesis, which allows one to create any individual. Many platforms have implemented protective safeguards that prevent the generation of high-profile or public figures. In the meantime, private individuals remain vulnerable to identity misappropriation, as they are often not protected by such systemic restrictions. Other ethical risks of high-fidelity synthesis extend to identity fraud and malicious manipulation of political discourse, particularly

through real-time avatar technologies such as Higgsfield AI that enable deceptive live video conferencing [10]. Furthermore, these tools facilitate online exploitation through the creation of non-consensual mature content and the systemic misleading of minors, posing severe threats to digital safety and individual privacy.

All of the threats mentioned above propose an urgent need for tools that can recognize synthetic content, such as detecting micro-expressions using a hierarchical transformer network [11].

6 Conclusions and future work

6.1 Discussion

The majority of the participants in this study were able to successfully identify the real person among the generated avatars. However, a subset of participants either selected an AI-generated avatar or expressed uncertainty in their decision. Some participants familiar with AI avatars were still misled, while some unfamiliar participants correctly identified the real video. This indicates that general AI literacy does not necessarily translate into detection ability. Preliminary results show that AI-generated avatar technology could be close to a point of reaching a level of realism sufficient to challenge human perception. Consequently, there is a need for the parallel development of effective AI-generated content detection technologies available to the general population.

6.2 Technical aspect

Among the eight evaluated tools, Flow by Google achieved the highest overall perceived naturalness rating (M=4.0), closely followed by Akool (M=3.43), making them the most convincing in terms of visual and vocal fidelity. These results were contrary to expectations, given that HeyGen was the most frequently recommended tool among online users and received the most positive reviews prior to evaluation. FalcoCut and AI Studios both use Kling as their underlying engine, however, these tools have been evaluated with lower scores than Akool, which also uses Kling. This suggests that the way a tool implements and customizes the API has a significant impact on the output quality. Contrary to expectations, three tools that offered custom voice generation (HeyGen, Akool, and AI Studios) did not consistently outperform prompt-based tools that generate audio automatically.

6.3 The Problem of the Uncanny Valley

Six of the eight AI avatar generation platforms evaluated triggered a sense of eeriness among participants, indicating that the uncanny valley is still a present problem in this field. The previously mentioned is a consequence of inadequate appearance of avatars, lacking naturalness of eye behavior or realism of facial expressions, which can be seen by ratings represented in Tables 3-11. In addition to looks, avatars lack vocal ability, as in the example in Table 11 where appearance received slightly better ratings. However, this may be attributed to limitations of the tools used, as the study deliberately focused on freely accessible platforms rather than premium or professional solutions.

6.4 Future work

Future research should include a larger sample of participants with greater demographic diversity, which would provide better insight into how the perception of AI avatars varies between different populations and generations. For a better understanding of the full capabilities of AI tools in this field, premium software should be considered. All of the above could be conducted at two different points over a multi-year period to analyze the ongoing development of artificial intelligence for content generation.

References

[1] Akool. Akool image to video: Ai-powered animation tool, 2026.

[2] Alibaba Qwen Team. Qwen chat, 2026.

[3] Jasmin Baake, Josephine Schmitt, and Julia Metag. Balancing realism and trust: Ai avatars in science communication. *Journal of Science Communication*, 24(2):A03, 2025.

[4] DeepBrain AI. Ai studios: Ai avatars, 2026.

[5] FalcoCut AI. Ai video generator, 2026.

[6] Google. Gemini: Multimodal ai architecture, 2026.

[7] Google Labs. Flow: An AI creative studio, 2026.

[8] Syed Ali Haider, Srinivasagam Prabha, Cesar Abraham Gomez-Cabello, Ariana Genovese, Bernardo Collaco, Nadia Wood, Mark A Lifson, Sanjay Bagaria, Cui Tao, and Antonio Jorge Forte. Artificial intelligence physician avatars for patient education: A pilot study. *Journal of Clinical Medicine*, 14(23):8595, 2025.

[9] HeyGen. Heygen AI avatar builder, 2026.

[10] Higgsfield AI. Higgsfield: Video creation for everyone, 2026.

[11] Md Ismail, Shah Md Labib Shadik, Md Sajid Ullah Sohan, MD Bhuiyan, and Lamia Khan Shoily. *An efficient approach for deepfake detection employing micro-expressions with a hierarchical transformer network*. PhD thesis, Brac University, 2025.

[12] Kling AI. Kling AI: Video Generation Platform, 2025.

[13] Mordor Intelligence. Digital human market - growth, trends, covid-19 impact, and forecasts (2026 - 2031). Market research report, Mordor Intelligence, 2026.

[14] Masahiro Mori, Karl F MacDorman, and Norri Kageki. The uncanny valley [from the field]. *IEEE Robotics & automation magazine*, 19(2):98–100, 2012.

[15] NVIDIA Corporation. Create digital humans with generative AI, 2026.

[16] Zhiyu Tao, Yanyan Liu, Junsheng Qiu, and Shengwei Li. Impact of virtual avatar appearance realism on perceptual interaction experience: a network meta-analysis. *Frontiers in Psychology*, 16:1624975, 2025.

[17] Vidnoz AI. Image to video ai, 2026.

[18] Wikipedia contributors. Diella (ai system), 2024.

[19] Ang Yang Yi and Tang Jia Huey. Ai avatars in action: A review of applications, challenges, and future potential. *Grenze International Journal of Engineering and Technology*, January 2025.