Perceiving Age Beyond Reality: Human and Machine Judgments of Virtual and AI-Generated Humans

Ana Carolina de Oliveira Xavier[1], **Victor Flávio de Andrade Araujo**[1,2]
Paulo Ricardo Knob[1,3], **Soraia Raupp Musse**[1,2,3]

[1]Escola Politécnica – PUCRS – Porto Alegre, Brazil

[2]Kunumi Institute, Brazil

[3]INCT SIMAI, CNPq, Brazil

a.xavier004@edu.pucrs.br, victor.araujo@edu.pucrs.br
paulo.Knob@edu.pucrs.br, soraia.musse@pucrs.br

Abstract. *This paper investigates how age is perceived and estimated in three categories of facial representations: Real Humans (photographs), Virtual Humans (3D MetaHuman renderings), and Digital Humans (photorealistic faces generated by a text-to-image model). We evaluate two age ranges (young: 20–30; older: 60–70) using a triangulated methodology: (i) four commercial age-estimation applications, (ii) a vision–language model (CLIP) under a zero-shot protocol, and (iii) an online user study ($N = 112$). Results indicate that commercial apps and CLIP exhibit their largest errors on 3D Virtual Humans (VHs), suggesting a domain generalization gap, while Artificial Intelligence (AI) generated 2D faces can be easier to classify than some real photographs. Human judgments show high accuracy for young AI-generated faces, but older AI-generated faces produce polarized responses. Overall, 3D VHs remain the most challenging domain for robust age perception.*

1. Introduction

The digital representation of humans is becoming increasingly ubiquitous, ranging from avatars in metaverses (Virtual Humans – VH) to photorealistic images generated by artificial intelligence (Digital Humans – DH). Tools such as Epic Games' MetaHuman Creator[1] enable the creation of high-fidelity virtual characters, illustrated in Figure 2 [Dalvi and Tuarob 2021], while generative models such as DALL-E (used via ChatGPT 4.0)[2] can produce faces (DHs), shown in Figure 4, that are at first glance indistinguishable from real photographs (Real Humans – RH). The ability to perceive and correctly estimate facial age has social, security (identity verification), and medical implications [Kumar et al. 2009]. The literature frequently reports the effectiveness of age and gender classification algorithms on real images [Levi and Hassner 2015, Rothe et al. 2015, Pan et al. 2018], the actual level of effectiveness remains questionable, particularly given the significant biases demonstrated by these systems. Studies such as "Gender Shades" [Buolamwini and Gebru 2018] and NIST reports [Grother et al. 2019] revealed demographic disparities in accuracy, motivating the creation of datasets such as FairFace [Karkkainen and Joo 2021] and Diversity in Faces [Merler et al. 2019] to mitigate these issues.

[1]Available at: https://www.unrealengine.com/en-US/metahuman
[2]Available at: https://openai.com/index/dall-e-3/

However, understanding age perception in synthetic humans is crucial for practical domains such as online safety and identity verification. If DHs can bypass automated age-gating systems (such as those deployed by Meta on Instagram to restrict underage access[3]), evaluating algorithm accuracy on synthetic faces becomes essential, as these systems have already been shown to be less accurate for older age groups[4]. This work addresses the following question: how do age detection algorithms and human perception interpret age in synthetic faces (VH and DH)? Existing literature primarily focuses either on age progression generation [Zhang et al. 2017, Antipov et al. 2017, Kim et al. 2020] or perception of real faces [Liu et al. 2015]. More recently, Vision-Language Models (VLMs) have begun to be investigated for bias analysis [Hausladen et al. 2025]. Based on this problem, the study was guided by the following hypotheses:

- **H1:** Age recognition applications will present greater difficulty (higher error margins) when estimating the age of Digital Humans (DH) and Virtual Humans (VH) compared to Real Humans (RH).
- **H2:** Human perception (user study) will be more accurate than software applications in age discrimination, regardless of human type.
- **H3:** The artificial application of ten skin tones to VH characters will introduce significant bias in age estimation compared to baseline characters, observable in both algorithms and human perception.
- **H4:** The CLIP (Contrastive Language-Image Pre-Training) VLM [Hausladen et al. 2025] will demonstrate more robust semantic understanding of age than traditional facial recognition applications.

2. Related Work

The literature supporting this work spans four interconnected areas: facial age estimation, bias studies in facial recognition, VH creation, and the datasets used to train these systems. The first area, age estimation and progression, focuses on research training Deep Learning models to estimate age from photographs [Levi and Hassner 2015, Rothe et al. 2015, Pan et al. 2018]. A complementary research direction explores facial aging generation using Generative Adversarial Networks (GANs) [Antipov et al. 2017, Zhang et al. 2017] or calibrating perceived age during face synthesis [Kim et al. 2020]. However, these algorithms remain limited by inherent biases. Seminal works such as "Gender Shades" [Buolamwini and Gebru 2018] and the NIST FRVT report [Grother et al. 2019] demonstrated significantly higher error rates for specific demographic groups, particularly dark-skinned women.

The third research area concerns virtual and synthetic humans. The pursuit of realistic VHs dates back decades [Thalmann and Musse 2000], yet the *Uncanny Valley* phenomenon — the discomfort reported when avatars appear almost human — remains a challenge [Dalvi and Tuarob 2021]. The emergence of high-fidelity VH and Artificial Intelligence (AI) generated DHs introduces new perceptual challenges not addressed in earlier literature. Finally, the performance of these models depends intrinsically on training data. Classical datasets such as CelebA [Liu et al. 2015], alongside early facial attribute classification work [Kumar et al. 2009], popularized attribute prediction tasks.

[3] Available at: https://about.fb.com/news/2022/06/new-ways-to-verify-age-on-instagram/
[4] Available at: https://www.futurebeeai.com/knowledge-hub/aging-face-recognition-accuracy

In response to demographic bias issues [Buolamwini and Gebru 2018], modern datasets such as *FairFace* [Karkkainen and Joo 2021] — explicitly balanced across race, gender, and age — and *Diversity in Faces* [Merler et al. 2019], designed to broaden global facial representation, were introduced. More recently, research has expanded to investigate how VLMs such as CLIP perceive social attributes [Hausladen et al. 2025], opening the analytical direction explored in this work.

3. Methodology

This study adopts a quantitative and comparative methodology divided into four main stages: image generation, computational analysis, human perception evaluation, and synthesis of results. Figure 1 illustrates the overall methodological workflow.

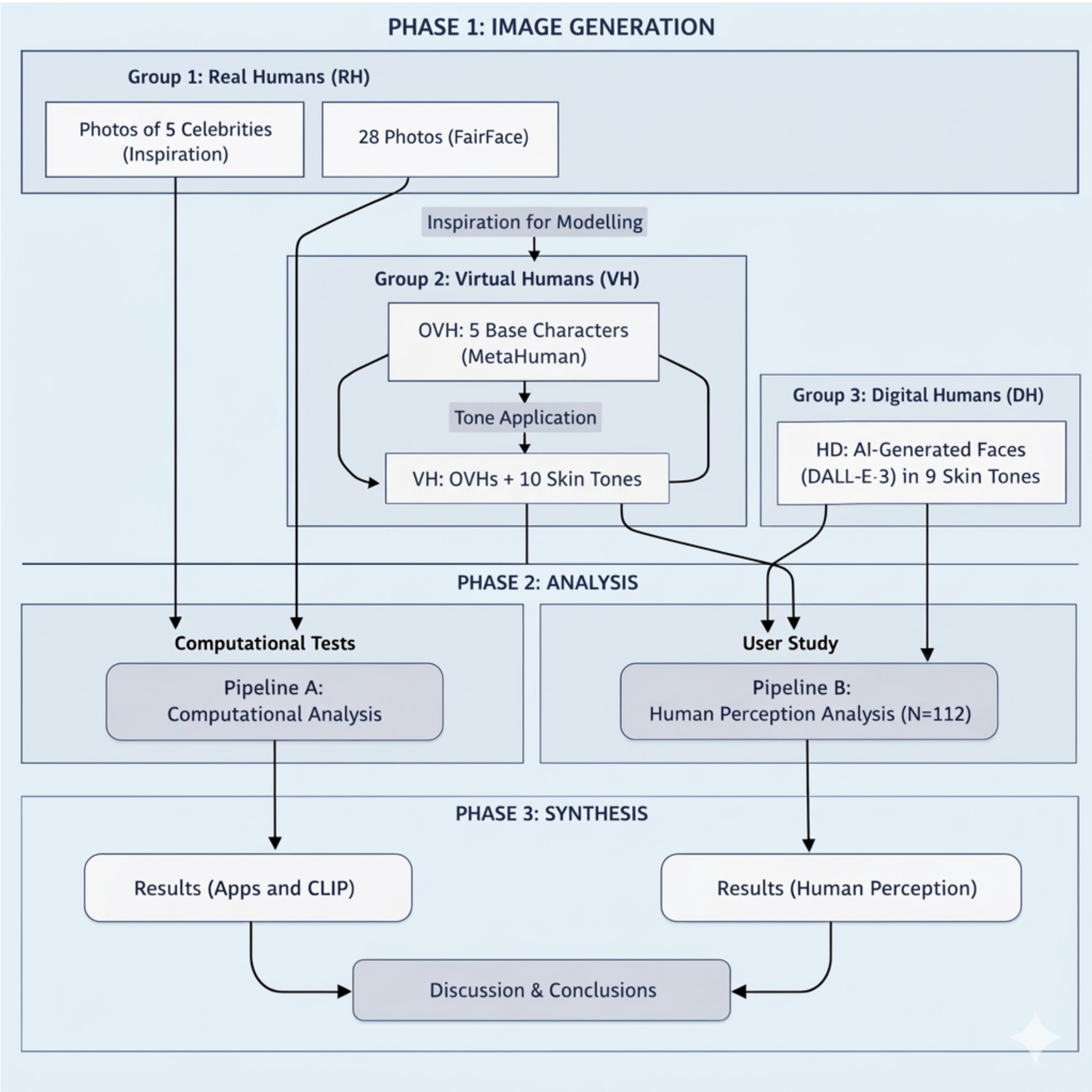


Figure 1. Overview of the method.

3.1. Phase 1: Image Generation

Three datasets were constructed, each representing a distinct category of human representation: **Group 1: Real Humans (RH) -** Images selected from the FairFace dataset[Karkkainen and Joo 2021] and publicly available photographs of celebrities, used as real-world baselines; **Group 2: Virtual Humans (VH) -** Characters created using MetaHuman Creator, ensuring controlled variation in age group and identity attributes; **Group 3: Digital Humans (DH) -** Photorealistic faces generated using DALL-E through structured textual prompts specifying age range, gender, and visual characteristics. All datasets were organized into two age groups: Young (20–30 years) and Elderly (60–70 years). Figures 2 and 4 illustrate some of the characters used in this work (respectively, VHs and DHs). Additionally, VHs were modified to include 10 distinct skin tones to analyze bias sensitivity across variations in appearance (Figure 3).

Figure 2. Example of Virtual Humans.

Figure 3. Example of HV dataset creation: the character "Neva" (young version) with the 10 skin tones applied.

3.1.1. Virtual Humans (VH)

The VHs group (a total of 110 images) was divided into two subsets, both created using the MetaHuman Creator software. First, five base characters, referred to as Original Virtual Humans (OVHs), were modeled to resemble the five selected celebrities ("Originals"). Following visual criteria commonly associated with facial aging (such as the appearance of wrinkles, changes in skin texture and roughness, and hair graying [Zhang et al. 2017, Kumar et al. 2009, Liu et al. 2015]) the software parameters were manually adjusted accordingly. For each reference character, two models were created to represent distinct age ranges: a young version (20-30 years) and an elderly version (60-70 years). Figure 2 illustrates two examples of base VHs shown in young and old versions, while Figure 3 presents a VH with different skin tones.

The aging parameters were adjusted to ensure visually realistic progression while preserving the VH's identity features. First, we used Epic Games' Metahuman Creator, which allows us to change parameters such as roughness, skin tone, etc. For the younger group (20-30 years old), the skin parameters were set with a texture of 0.2420 and roughness of 0.10, while the contrast, grayscale tone, and grayscale brightness (for both hair and beard) were kept at 0.0. In contrast, the configuration for the elderly group (60-70 years) involved more pronounced adjustments. Skin texture was increased to 0.8484, with contrast and roughness both set at 0.70. For hair, roughness was set at 0.45, with gray tone and brightness at 0.75. Eyebrows and facial hair were adjusted to a gray tone of 0.70 and a brightness of 0.50. We used three justifications for choosing these parameters: **Skin Texture and Roughness** were increased to simulate aging indicators such as wrinkles, loss of elasticity, and dermal irregularities [Zhang et al. 2017]; **Contrast** was enhanced to emphasize expression lines and facial depth, supporting visual perception of age [Kumar et al. 2009]; and **Grayness and Shine** were progressively adjusted in hair and facial hair to mimic the typical loss of pigmentation associated with aging [Kumar et al. 2009].

After creating the five OVHs with both age settings (young and old), each character received a base skin tone defined by specific texture coordinates (U, V) available in the Metahuman Creator system. The original tones were assigned as follows: Henry (Asian man, U=0.47, V=0.37), Hadley (white woman, U=0.35, V=0.63), Myles (white man 1, U=0.41, V=0.30), Wallace (white man 2, U=0.36, V=0.43), and Neva (black woman, U=0.79, V=0.61). All these OVHs and their names are standard options provided by the Metahuman Creator.

3.1.2. Digital Humans (DH)

A total of 54 images were generated using the DALL-E 3 image generation model (via ChatGPT-4). The stimulus strategy focused on high realism, neutral backgrounds, and variations in age and skin tone. To ensure consistency across the dataset, the following stimulus structure was employed: "Consider a scale of 1 to 10, where 1 represents the lightest skin tone and 10 the darkest. Generate a photo of an [woman/male/asian male] aged between [60-70 / 20-30] years. The image should be a close-up of the face, similar to a passport photo (10x10 cm). Apply each skin tone from the scale to this person, generating 10 identical images of this [Asian man/woman/male], but with 10 different skin tones. The faces should be fictional and realistic, with a neutral background". The DHs were created based on three profiles (male, female, asian male) for both age ranges. Figure 4 shows examples of DHs, young and elderly women, and young and elderly Asian male. Due to limitations of the model during the iterative process, a total of 9 distinct variations per character were generated instead of the 10 requested, resulting in the final set of 54 images (3 profiles × 2 age ranges × 9 tones).

Figure 4. Examples of AI Digital Humans

3.2. Phase 2: Analysis

3.2.1. Computational Analysis

We used two computational approaches to evaluate our images. The first method involved the use of four commercial age-estimation applications available on the Google Play Store, for example, Face Age Detector[5], Face scan[6], Check my age[7], and Age Detector[8]. For each sample, the predicted ages were recorded and compared against the ground-truth age group. We calculated the Mean Absolute Error (MAE), which defined as the difference between the target age average and the predicted age, as well as the accuracy based on age range criteria. This analysis allowed us to assess how effectively commercial tools, typically trained on real photographs, generalize when applied to synthetic representations. In the second method, we used CLIP [Radford et al. 2021], a neural network trained with several pairs (image, text). In this case, we used the zero-shot approach, in which CLIP classified each face into age categories (e.g., "young" vs."old") using text clues, selecting the class with the highest similarity. The predicted category corresponded to the prompt with the highest similarity score (probability - a value ranging from 0 to 1). Accuracy was computed per dataset type (RH, VH, DH).

3.2.2. Human Perception Analysis

We conduct an online survey ($N = 112$) with volunteer participants. Firstly, they answered the ethics-approved consent form[9], volunteers gave demographics (age, gender, education, Computer Graphics (CG) familiarity). After that, they viewed a random subset of synthetic faces (VH and DH variants). They were asked to categorize the perceived age of the faces into age ranges of ten to ten years, ranging from '10-19' to '+100'. This task was designed to evaluate the effectiveness of age-related design cues (such as skin texture and facial structure) within the synthetic models. Responses were subsequently

[5]https://play.google.com/store/apps/details?id=com.face.scanner.age.calculator.detector
[6]https://play.google.com/store/apps/details?id=com.tikamori.face.age.recognition
[7]https://play.google.com/store/apps/details?id=com.neurotec.checkmyage
[8]https://play.google.com/store/apps/details?id=com.detector.age
[9]Project "Assessments of Human Perception in Virtual Characters and Crowds", number 46571721.6.0000.5336, approved by the Ethics Committee of the Pontifical Catholic University of Rio Grande do Sul.

classified as correct, underestimated, or overestimated relative to the character's ground-truth age, allowing us to quantify how visual aesthetics influence the observer's judgment of maturity and life stage.

4. Results - Phase 3: Synthesis

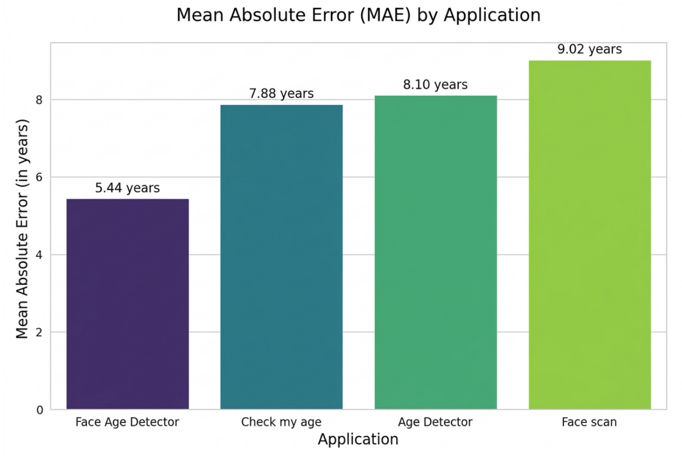
Firstly, the demographic data showed that the vast majority of participants (over 70%) were in the 18-24 and 25-35 age groups, more than 60% identified as women, the majority possessed a high school diploma or bachelor's degree (65%), and CG familiarity 81.2%.

4.1. Commercial Apps

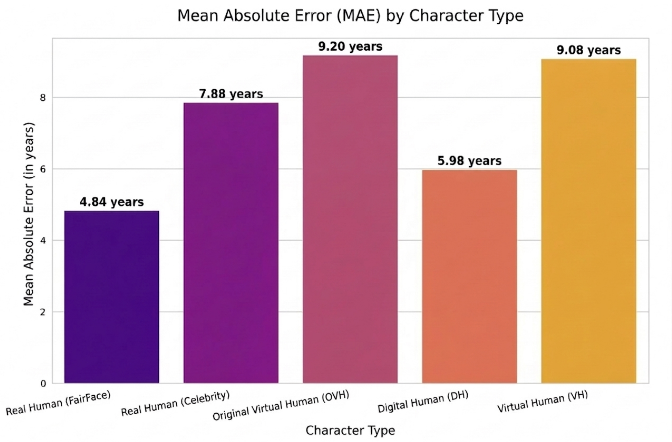
Across all groups, application performance varied substantially, but a consistent hierarchy emerged: real faces (FairFace) exhibited the lowest Mean Absolute Error (MAE), while 3D faces based on MetaHuman exhibited the highest MAE. Notably, AI-generated 2D digital faces can achieve a lower MAE than some real celebrity photos, suggesting they artificial but realistic, i.e., the faces are realistic, but the images lack real-world noise like motion blur or complex lighting. We highlighted the aggregate comparison in Figure 5. First, the MAE was evaluated, representing the average difference in years between the target age and the estimated age. Figure 5 (a) showed that *Face Age Detector* was the most accurate (MAE of 5.44 years), while *Face scan* was the least accurate (MAE of **9.02 years**). The apps *Check my age* (MAE of 7.88 years) and *Age Detector* (MAE of 8.10 years) showed intermediate performance. The analysis of the aggregated MAE by character type (Figure 5 (b)) revealed that the FairFace RHs group had the lowest error (MAE of **4.84 years**), while the DHs had the second best performance (MAE of **5.98 years**), and the Celebrity RHs had an MAE of **7.88 years**. The two VHs groups showed the worst performance, with the VHs group having an MAE of 9.08 years, and the OVHs having the highest error of all, at 9.20 years. We also conducted an analysis looking at the young (20-30 years) and elderly (60-70 years) age groups. In the young age group, only the Face Age Detector and Face Scan apps in the VHs category had accuracy rates lower than 50%, with the former close to 50% and the latter below 20%. However, the performance of all apps dropped drastically when analyzing the elderly age group. Only Face Age Detector in the FairFace (RHs) and DHs categories, and Check My Age had performance close to 60% as the highest values among all. All others had percentages lower than 40%.

4.2. CLIP Zero-Shot Classification

Regarding CLIP, we used two prompts to measure the accuracy percentages in relation to the two age ranges we had. So we use the prompt "a photo of a young person" for the 20-30 age range, and the prompt "a photo of an old person" for the 60-70 age range. For the group of RHs, composed of the FairFace dataset and celebrity photos, CLIP showed high accuracy rates. The model achieved 83.33% accuracy in age classification. Regarding DHs, CLIP also achieved a high accuracy rate of 96.3% for age (higher than for RHs). This may suggest that CLIP is able to understand the dynamics of 2D images generated by artificial intelligence. Unlike the other two cases (RHs and DHs), CLIP showed a low percentage in relation to VHs (23.6%). This indicates that CLIP may have greater difficulty in interpreting the signs of aging (age) present in 3D model renderings, which may be a challenge for the use of visual language models in 3D rendered images.



(a) MAE by Commercial Application



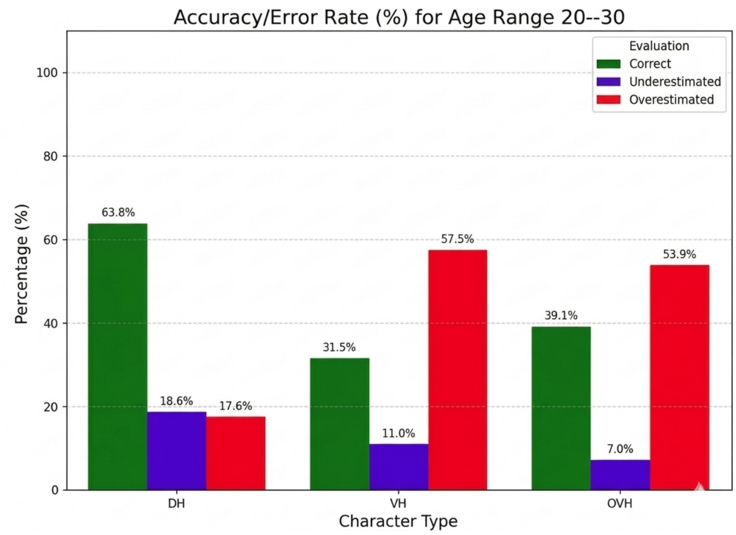
(b) MAE by Character Type

Figure 5. Mean absolute error (MAE) grouped by Commercial Application (a), and by Character Type (b).

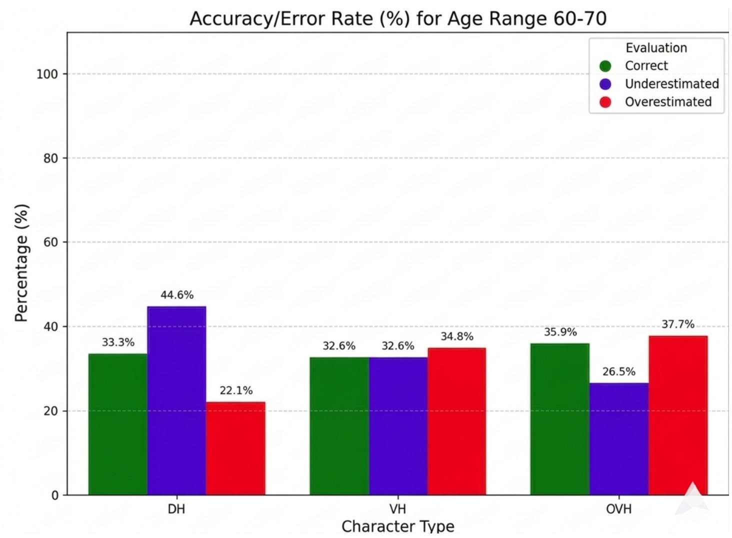
4.3. Human Perception Analysis

Human judgments demonstrate high accuracy for young DHs, whereas older DHs lead to response polarization. In this subset, participants were split between underestimation, correct estimation, and overestimation, suggesting ambiguous age cues. Comparison of the summary plots (Figure 6) reveals that DHs achieved the highest overall "Correct" percentage, followed by OVHs, and then VHs. Furthermore, VHs exhibited higher percentages of both "Underestimation" and "Overestimation" compared to OVHs. These results suggest that the process of applying different skin tones to the base models (OVHs) may have diminished the perceived realism of age-related features, leading to lower consensus in the VH category. We also analyzed the percentage of age perception in different age ranges (10-19 to 100+, referring to the response options of the applied questionnaire). Regarding young characters, most responses for all three groups were concentrated in the age range of the 20-29 and 30-39 options. Specifically, in the first of these options, DH characters achieved the highest percentage of choice (approximately 64%), followed by OVH (40%), and VH (32%). Interestingly, for the 30-39 age range, DH characters had a sharp drop in responses (below 20%), having a lower value than VHs (35%) and OVHs

(45%). Regarding the elderly characters, the results showed that the highest concentration of responses was within the correct age range of 60 to 69 years. Specifically, the OVHs achieved the highest accuracy in this range (approximately 36%), followed by the DHs (around 33%) and VHs (approximately 32%). The data distribution showed an almost Gaussian pattern, with a steady increase in percentages up to the 60-69 age range and a gradual decline in subsequent older age ranges. For example, for the 70-79 age range, there was a drop in consensus for DH characters (falling to approximately 17%) compared to OVHs and VHs characters, which remained above 20%. In general (young and old cases), we can note that people's responses were in the direction of the age group that correctly represented the age of both the HVs and the DHs.



(a) Human perception outcomes for young faces (20–30)



(b) Human perception outcomes for older faces (60–70)

Figure 6. Correct vs. Under/Overestimation by Character Type.

5. Discussion

This section aims to discuss our results in relation to our hypotheses. First of all, H1 was partially confirmed, as MAE metrics (Fig. 5) revealed performance gaps across digital representations. The MAE for Virtual Humans (OVH: 9.20; VH: 9.08 years) nearly doubled that of the FairFace RH group (4.84 years), likely due to a domain shift where 3D-rendered shaders deviate from the real distributions typical of CNN training datasets [Levi and Hassner 2015, Rothe et al. 2015, Buolamwini and Gebru 2018]. Conversely, DH results challenged H1; their MAE (5.98 years) outperformed the Celebrity RH group (7.88 years). This suggests that “cleaner” 2D generative images (lacking real-world noise like motion blur or complex lighting) provide more accessible inputs for age estimators than “in-the-wild” photography. Finally, the performance decline in the elderly cohort (60-70 years) across all categories underscores a systemic bias toward younger, more standardized facial features.

In H2, the hypothesis that human perception would be more accurate than software applications in age discrimination across all types of characters was confirmed. The results indicated that human observers possess a certain capacity to interpret age-related visual cues in synthetic faces, surpassing commercial applications. The disparity in performance is most evident when comparing maximum accuracy rates, where in the young adult group (20-30 years old), the most accurate app (Facial Age Detector) achieved an accuracy of approximately 55%. In contrast, participants obtained a higher success rate, particularly for the DHs, where 64% of responses correctly identified the 20-29 age range. Even for the older adult group (60-70 years old), where the app’s accuracy dropped to around 30% (Age Detector), human judgment remained more robust, with the OVHs achieving a correct estimation rate of 36%. This suggests that humans may be able to synthesize ambiguous age cues, such as the interaction between skin texture and lighting, more effectively than current CNN-based models.

Hypothesis H3, which posited that the artificial application of ten skin tones to VH characters would introduce a significant bias in age estimation, was partially confirmed, revealing a clear divergence between human and algorithmic perception. Regarding automated estimators, the hypothesis was largely inconclusive; the MAE for the base models (OVH: 9.20 years) and for those with varying skin tones (VH: 9.08 years) remained virtually identical. This may suggest that the main bottleneck for these applications is not chromatic variation, but the inherent nature of 3D rendered shaders, which pose a persistent challenge for models trained with organic data. On the other hand, H3 was confirmed by human perception. Participants demonstrated a higher percentage of correct predictions for the OVH models compared to the VH models. This may indicate that the artificial application of skin tones likely created confusion regarding signs of aging (such as micro-wrinkles and skin irregularities), diminishing the perceived realism of the characters and leading human observers to classify them as younger than their actual age.

Hypothesis H4, which postulated that CLIP would exhibit a more robust semantic understanding of age compared to traditional facial recognition applications, was partially confirmed, demonstrating superior performance primarily in 2D image domains. CLIP outperformed traditional estimators in the RH and DH categories. This accuracy may suggest that CLIP’s semantic architecture is capable of decoding the aging dynamics present in AI-generated 2D images, overcoming the limitations of traditional applications that struggled with these types of features. However, for VHs, CLIP's accuracy decreased, indicating that even advanced vision-language models [Hausladen et al. 2025] may encounter difficulties interpreting artificial aging signals in 3D rendered facial shaders. These results show that a gap may exist in how current vision-language models process 3D rendered 2D facial features versus naturalistic or AI-generated 2D facial features, aligning with the "Uncanny Valley" discussion by [Dalvi and Tuarob 2021].

5.1. Limitations and Future Work

Limitations include a restricted number of base identities for virtual and digital sets, and a participant demographic skewed toward young adults. Additionally, the use of generated images may have introduced variability that was not fully controlled or quantified (in other words, controlling models without free apps), and we also did not perform statistical significance tests, which could be done in more controlled studies in the future. Future work should expand identities, include intermediate age ranges, diversify participant demographics, and study other facial factors (expression, lighting, style). A controlled real-human subset within the human study would also allow direct baseline comparisons.

6. Final Considerations

This study analyzed age perception in RH, VH, and DH faces by triangulating results from commercial age recognition applications, a VLM (CLIP), and a human perception study. Our main results revealed that commercial applications, trained primarily on real-world datasets, can fail when confronted with synthetic images. Specifically, these models failed to accurately process 3D-rendered characters (OVH and VH). Paradoxically, both algorithms and human observers found 2D faces (DH) the easiest to classify among the synthetic groups, suggesting that those generative images may align more closely with prototypical aging patterns than complex 3D shaders. Future research should aim to expand the VHs and DHs datasets to include a wider range of baseline characters and intermediate age groups. Investigating the influence of other facial attributes, such as expressions and lighting conditions, on the synthetic perception of age would provide a deeper understanding. Finally, replicating the study with humans using a more diverse and stratified demographic sample, including a direct baseline with RHs, will be essential to more accurately establish the limits of age perception in the ever-evolving landscape of DHs and VHs.

7. Acknowledgements

This study was partly financed by: the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior – Brazil (CAPES) – Finance Code 001; INCT SIMAI, CNPq #408330/2024-4; Kunumi Institute - The authors thank the institution for its financial support and commitment to advancing scientific research.

References

Antipov, G., Baccouche, M., and Dugelay, J.-L. (2017). Face aging with conditional generative adversarial networks. *arXiv preprint arXiv:1702.01983*.

Buolamwini, J. and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91. PMLR.

Dalvi, C. and Tuarob, S. (2021). A survey on the uncanny valley in virtual humans. *IEEE Access*, 9:53569–53587.

Grother, P., Ngan, M., and Hanaoka, K. (2019). Face recognition vendor test (frvt) part 3: Demographic effects. Technical Report NISTIR 8280, National Institute of Standards and Technology.

Hausladen, C. I., Knott, M., Camerer, C. F., and Perona, P. (2025). Social perception of faces in a vision-language model. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, pages 639–659.

Karkkainen, K. and Joo, J. (2021). Fairface: Face attribute dataset for balanced race, gender, and age. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1548–1558.

Kim, T.-H., Kwon, H.-J., Choi, J.-Y., and Kweon, I. S. (2020). Face synthesis with identity preservation and perceived age calibration. In *European Conference on Computer Vision (ECCV)*, pages 594–610. Springer.

Kumar, N., Berg, A. C., Belhumeur, P. N., and Nayar, S. K. (2009). Attribute and simile classifiers for face verification. In *2009 IEEE 12th International Conference on Computer Vision*, pages 365–372. IEEE.

Levi, G. and Hassner, T. (2015). Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 34–42.

Liu, Z., Luo, P., Wang, X., and Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of the IEEE international conference on computer vision*, pages 3730–3738.

Merler, M., Ratha, N., Ferrara, M., Zhang, J., and Smith, J. R. (2019). Diversity in faces. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 0–1. IEEE.

Pan, X., Han, Y., Shan, S., and Chen, X. (2018). Mean-variance loss for deep age estimation from a face. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5285–5294.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR.

Rothe, R., Timofte, R., and Van Gool, L. (2015). Dex: Deep expectation of apparent age from a single image. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 10–15.

Thalmann, D. and Musse, S. R. (2000). The foundations to build a virtual human: The six-layer approach. In *Proceedings of the workshop on virtual environments*, pages 11–19.

Zhang, Z., Song, Y., and Qi, H. (2017). Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5810–5818.