Human and AI Face Perception Experiment

Lazer Yerkiman
Educational Foundation "IQANAT"
Almaty, Kazakhstan
Email: erkimanlazer@gmail.com

***Abstract*—The Face Perception Experiment explores how accurately both humans and artificial intelligence can assess key facial traits, such as age, gender, and ethnicity, from images. For this study, 15-year-old participants were invited to evaluate facial characteristics from a set of images. Then, their assessments were compared to the predictions made by DeepFace, an advanced AI-powered facial recognition system. The results indicate that human participants outperform DeepFace in age estimation and achieve higher accuracy in gender classification.**

I. Introduction

Facial perception is a fundamental aspect of human cognition, allowing individuals to quickly infer attributes such as age, gender, and racial background based on visual cues. While humans rely on instinct, experience, and cultural knowledge, artificial intelligence models leverage deep learning algorithms trained on large datasets to recognize patterns and predict facial attributes.

In this experiment, a group of 15-year-old participants were asked to estimate the face attributes of individuals from 22 different face images. Their responses were compared with predictions from DeepFace [1]. The study assesses which aspects of facial perception AI outperforms humans in and where humans demonstrate better accuracy.

II. Methodology

This study employs a comparative experimental design to analyze and contrast the performance of human perception and AI-based facial recognition models. The experiment consists of two primary components.

A. Human-Based Face Perception Survey

A group of 15-year-old participants (n = 20) was asked to predict the age, gender, and race of the individuals in 22 facial images. The images featured diverse individuals to ensure a balanced representation of different age groups, genders, and racial categories. Participants independently made their decisions through an online survey. Responses were stored in a structured dataset. The binary correctness metric (1 = correct, 0 = incorrect) was applied to each prediction. After that, aggregated human accuracy scores were calculated for each attribute.

B. AI-Based Face Attribute Prediction

We employed DeepFace, a deep learning-based facial analysis model for the same 22 facial images evaluated by human participants. The model predicts age (continuous value), gender (binary classification), and race (multiclass classification) [2]. Prediction Pipeline:

- Age Prediction $\rightarrow$ Converted into categorical bins (e.g., 13-19, 20-29).
- Gender Prediction $\rightarrow$ Classified as "Man" or "Woman".
- Race Prediction $\rightarrow$ One of six categories: "White," "Black," "Asian," "Indian," "Middle Eastern," "Latino Hispanic".

To quantitatively compare human and AI performance, we employed the following classification and regression metrics:

1) **Accuracy**: The proportion of correctly classified cases across all categories.
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$
Where:
   - TP = True Positives (correct predictions)
   - TN = True Negatives (correct non-predictions)
   - FP = False Positives (incorrect predictions)
   - FN = False Negatives (missed correct cases)
2) **Precision (Positive Predictive Value)**: The ratio of correctly predicted attributes to the total predicted attributes.
$$Precision = \frac{TP}{TP + FP} \tag{2}$$
Higher precision indicates that the model is making fewer false positive errors.
3) **Recall (Sensitivity)**: The ratio of correctly classified attributes to the total actual instances of that attribute.
$$Recall = \frac{TP}{TP + FN} \tag{3}$$
Higher recall means fewer false negatives (missed correct cases).
4) **Mean Absolute Error (MAE) – Regression Metric:** Since age prediction is a numerical task, we also compute MAE to measure how far predictions deviate from actual values.
$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{4}$$
Where:
   - y_i = True value (actual age)
   - $\hat{y}_i$ = Predicted value (estimated age)
   - N = Number of samples

III. Results

The Comparison of Human and AI Accuracy Fig. 1 highlights the differences in performance between human participants and AI predictions. The analysis shows that humans are better at estimating ages, potentially due to contextual understanding. Whereas AI models struggle with age-related variations. In Gender Prediction, human accuracy reached 99.17%, surpassing AI's 86.36%. This is expected since gender classification is often visually apparent to humans.

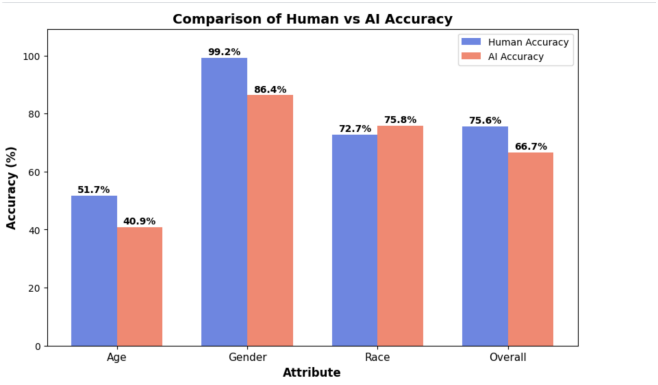

Fig. 1. Comparison of Human and AI Accuracy

TABLE I
Performance Metrics for AI and Human Accuracy

Metric	DeepFace Score	Human Score
Gender Precision	89.09%	99.17%
Gender Recall	86.36%	99.17%
Gender MAE	0.136	0.833
Age Precision	40.91%	51.67%
Age Recall	40.91%	51.67%
Age MAE	1.409	48.33
Race Precision	81.06%	75.83%
Race Recall	72.73%	75.83%
Race MAE	0.636	24.17

AI performed slightly better in race classification with a race precision of 81.06%, while human precision was 75.83% (Table 1). This might indicate that AI has a more systematic feature-based approach in identifying racial traits. Human cultural and cognitive biases might influence their judgment. The Mean Absolute Error (MAE) values in Table 1 show that AI had lower errors in both Age and Race classification, meaning that even when AI made incorrect predictions, the misclassification was less extreme than human errors

Fig. 2 presents a selection of images along with their true labels and predicted attributes from the DeepFace model. Misclassifications are particularly noticeable in age predictions, where AI tends to underestimate or overestimate certain age groups.

IV. Conclusion

This experiment demonstrates that while AI continues to improve, it falls short of human perception in key areas. Humans outperform AI in gender and age classification, whereas AI performs better in race classification but with potential biases. The study emphasizes the importance of refining AI training methodologies to improve facial perception capabilities.

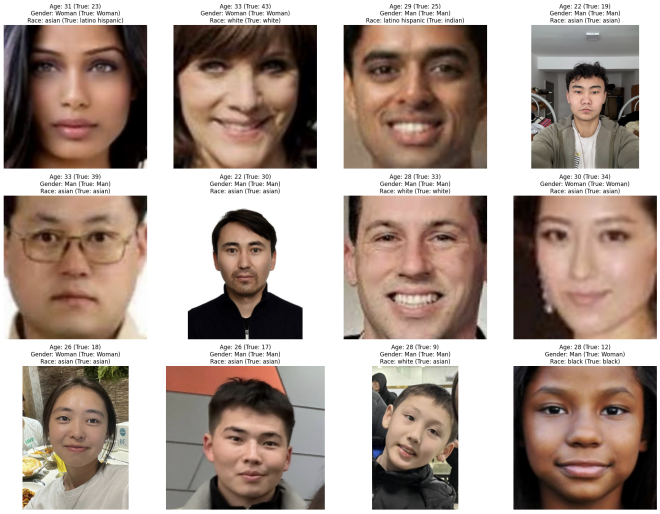

Fig. 2. DeepFace predicted results with true labels

References

[1] S. I. Serengil, "Deepface: A facial recognition and attribute analysis library for python," 2023, accessed: 1-Mar-2025. [Online]. Available: https://github.com/serengil/deepface

[2] E. Ardelia, J. Thenando, A. A. S. Gunawan, and M. E. Syahputra, "Predicting age and gender across different races using convolutional neural networks: A deep learning approach," in *International Conference on Information Technology and Computer Modeling (ICITCOM)*, 2024.