dEEG-fake: Experimental Design

Neurocognitively-guided modelling of virtual humans for enhanced realism in immersive media

Aileen Brama, Birgit Nierula, Florian Barthel, Arian Beckmann, Tilmann Stephani, Yonghao Chen, Anna Hilsmann, Vadim Nikulin, Michael Gaebler, Felix Klotzsche, Peter Eisert, Arno Villringer, Thomas Wiegand, Sebastian Bosse

Fraunhofer Heinrich-Hertz-Institute, Vision and Imaging Technologies Department, Interactive & Cognitive Systems Group
Einsteinufer 37, 10587 Berlin

Max-Planck-Institute for Human Cognitive and Brain Sciences, Stephanstraße 1A, 04103 Leipzig, Germany

Background

- Deepfake technology enables the creation of highly realistic manipulated media. Detection models often struggle to maintain accuracy when faced with diverse manipulation techniques (Wang et al., 2023).
- An alternative approach is to use EEG responses from human observers to capture the differences in perception between real and manipulated faces (Beckmann et al., 2024).
- Building on preliminary findings by Beckmann et al. (2024) this study investigates the perceptual quality in GAN-generated images and compares them to standard visual metrics, supporting the integration of neurophysiological data into deepfake detection.

Our goals:

1. Validate on FID as a perceptual quality metric.
2. Use EEG signals to inform realism assessment of GAN images.

Methods

- within-subject design
- dependent variable: realness
- independent variable: ERP amplitude
- binary classification (real vs. fake)

1) Behavioral pre-experiment:

- **Participants**: N = 10
- **Duration**: ~30min
- **Design**: 5 faces across 100 conditions= 500 stimuli
- **Timing**: 350ms stimulus + jittered breaks
- **Goal**: Identify perceptual threshold ≈50% real/fake

2) EEG experiment:

- **Participants**: N ≥ 2
- **EEG Setup**: 64 channels
- **Duration**: ~4.5 hours
- **Response** in 12.5% of trials via keypress (≤ 2s)
- **Design**:160 blocks × 100 trials = 16.000 stimuli
  - 4 experimental conditions
- **Timing**: 350ms stimulus + jittered breaks

All stimuli are synthetic face images generated via Generative Adversarial Networks (GAN). We use the Fréchet Inception Distance (FID) to quantify visual image quality enabling comparison between objective metrics and subjective perception.

Stimulus generation

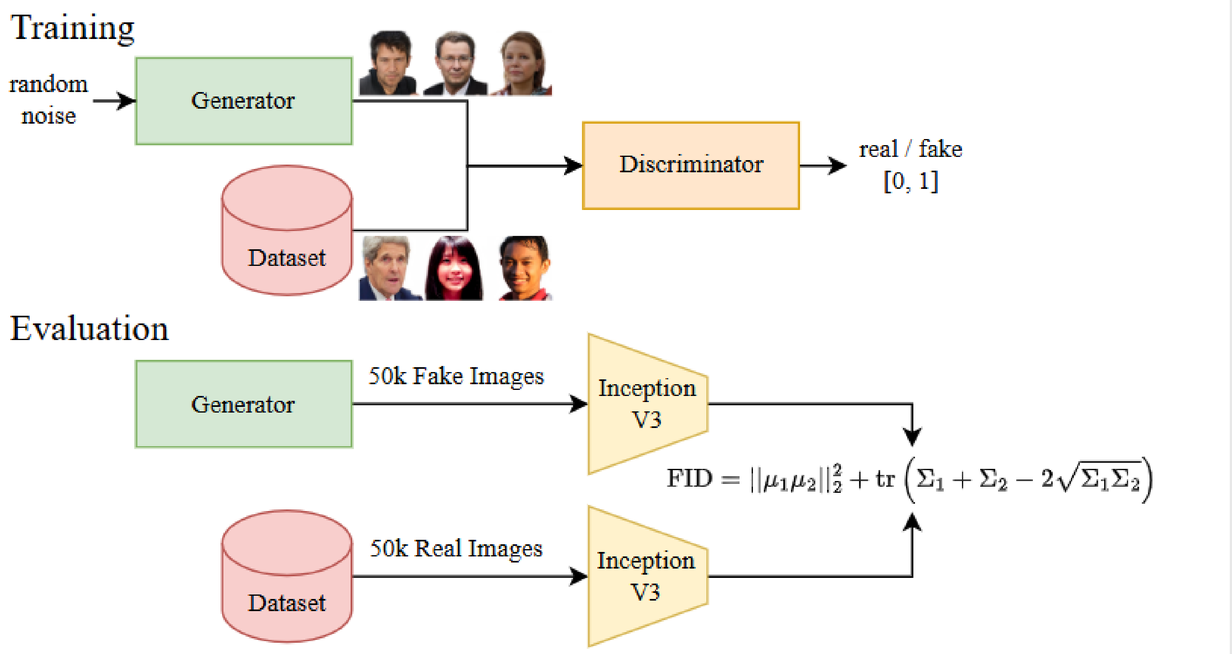

Figure 1: Overview of GAN based image generation where two neural networks the generator and the discriminator compete (Barthel et al., 2025). The generator creates synthetic data, while the discriminator tries to distinguish it from real data (Goodfellow et al., 2014). Evaluation is based on using the FID as a metric for assessing the visual quality of images. It compares from a pre-trained Inception-v3 network between real and generated images by their means and covariances (Heusel et al., 2017).

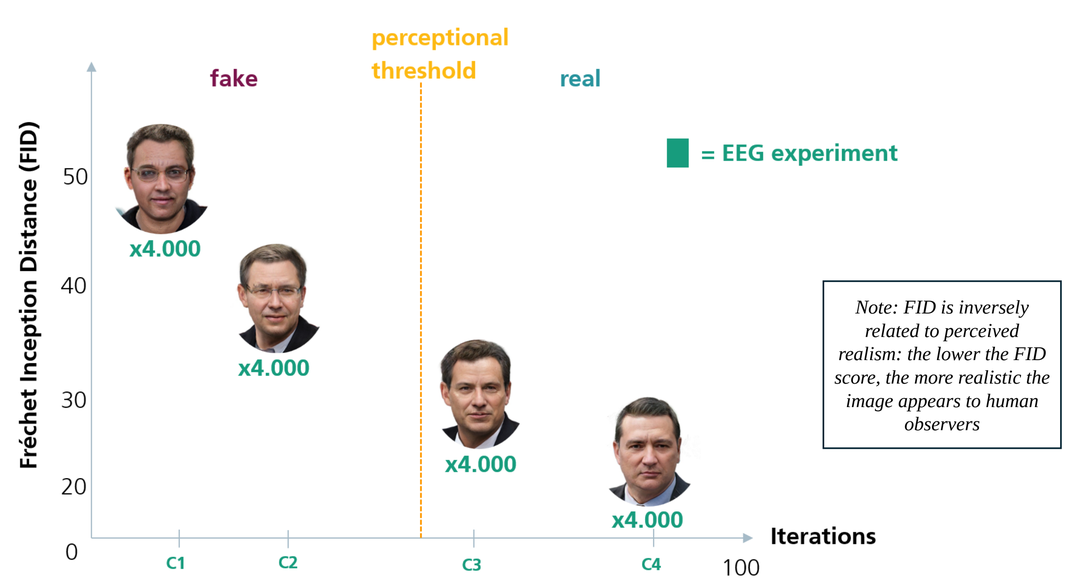

Figure 2: EEG experiment design across four stimuli conditions (C1–C4) categorized by increasing visual realism based on Fréchet Inception Distance (FID). The perceptual threshold was approximated in a preliminary pre-experiment. Images to the left of this threshold are classified as 'fake', while those to the right are perceived as more 'real'.

Existing systems optimize for statistical similarity but remain unaware of the perceptual cues that define realism for human observers.
By integrating EEG-based neurophysiological feedback we aim to close this foundational gap and guide image generation toward outputs that are not only high-quality, but perceptually convincing to human observers (Chen et al., 2023; Moshel et al., 2022).

Discussion

We expect that integrating neurophysiological feedback into AI training or evaluation processes can lead to more human-aligned Deepfake detection systems that better reflect human realism judgments.

Acknowledgements

This research is part of the NeuroHum project funded by the cooperation between the Max Planck Society and the Fraunhofer Society.

References

- Barthel et al., 2025, arXiv:2505.17590
- Beckmann et al., 2024, arXiv preprint, arXiv:2405.08527
- Chen et al., 2023 preprint, ResearchGate: 373355677
- Goodfellow et al., 2014, Advances in Neural Information Processing Systems, 27, 2672–2680
- Heusel et al., 2017, Advances in Neural Information Processing Systems, 30, 6626–6637
- Moshel et al., 2022 Vision Research, 198, 108085
- Wang et al., 2023 ACM Computing Surveys, 56(4), Article 85