scientific reports

Check for updates

OPEN

Neural signatures of hyper-realistic AI-generated faces: dissociating behavioral indistinguishability from implicit neural evaluation

Alice Mado Proverbio[1,2]✉ & Mariia Dosaikina[1,2]

Recent advances in generative artificial intelligence (AI) have produced hyper-realistic synthetic faces that are increasingly difficult to distinguish from real human faces, raising critical questions about how such stimuli are encoded by human brain. Behavioral studies indicate that observers frequently misclassify AI-generated faces as real and often judge them as more familiar/ attractive than genuine faces (hyperrealism). We investigated whether neural processing differentiates real from highly realistic AI-generated faces despite observers' limited behavioral discrimination ability. Thirty participants viewed 440 real and GAN-generated male/female faces while their EEG was recorded. Behavioral validation confirmed that AI-generated faces were poorly identified as artificial and were perceived as more familiar and aesthetically appealing. Face-evoked ERPs showed systematic modulation by realism despite task irrelevance. AI-generated faces elicited enhanced N250, P300, PN400, and late positivity; a reduced engagement of ventral temporal, parietal, and limbic networks was found for real faces, especially male ones, according to swLORETA. Although hyper-realistic faces surpass behavioral detection thresholds, the brain remains sensitive to their artificial origin. Neural markers of familiarity and aesthetic appraisal are enhanced for AI faces, likely reflecting algorithmic averaging that accentuates prototypical features and attenuates sexual dimorphism, thereby revealing a dissociation between explicit recognition and implicit neural evaluation.

Keywords Artificial intelligence, GAN-generated faces, Face perception, Event-related potentials (ERP), Neural familiarity, Aesthetic evaluation, Sex differences, Source reconstruction

For many years, virtual faces created using computer graphics have been used in research, making it important to understand how people respond to them as their prevalence in everyday life increases[1]. Studies have documented differences in responses to virtual versus real faces[2,3], including a reduced other-race effect, poorer face memory[4], and weaker neural responses[1]. More recently, advances in AI, particularly Generative Adversarial Networks (GANs), have enabled the creation of increasingly realistic artificial faces[5]. GANs, which are trained on photographs of real people to generate realistic images of entirely new faces[6], have made it progressively more difficult to distinguish real from artificial faces, with important real-world consequences such as deepfakes and social media scams[7].

Recent advances in generative models have substantially blurred the perceptual boundary between real and artificial human faces. Behavioral studies consistently show that observers struggle to reliably distinguish AI-generated faces from real ones, often performing near chance level, even when explicitly instructed to detect artificiality. Strikingly, several reports indicate that synthetic faces are sometimes judged as more trustworthy, more realistic[8,9] or socially appealing than real faces, suggesting that contemporary generative systems may exploit perceptual heuristics related to prototypicality and familiarity (e.g., Nightingale and Farid[10]). Together, these findings challenge long-standing assumptions about visual realism and raise critical questions about how artificial faces are represented in the human cognitive system.

Regarding electrophysiological evidences, Chen et al.,[11] demonstrated that the perceived realness of highly similar stylized faces could be decoded using steady-state visual evoked potential analyses. More broadly, exposure to highly human-like artificial stimuli elicits an uncanny valley response, with multivariate neural

[1]Cognitive Electrophysiology Laboratory, Department of Psychology, University of Milano-Bicocca, Piazza dell'Ateneo nuovo 1, 20162 Milan, Italy. [2]Master's Program in "Human-Centered Artificial Intelligence" HCAI, Department of Philosophy, University of Milan, 20122 Milan, Italy. ✉email: alice.proverbio@unimi.it; mado.proverbio@unimib.it

analyses revealing both early (100–200 ms) and late (~ 600 ms) neural signatures that reliably differentiate androids and realistic masks from real human faces[12]. Furthermore, Tarchi et al.[13] showed that although highly human-like artificial faces were behaviorally difficult to distinguish from real faces and elicited minimal neural differences, multivariate neural patterns applied to EEG spectral analysis revealed early perceptual and late top-down signatures consistent with the uncanny valley response. Again, Wheatley et al.[14] reported evidence of neural markers supporting a distinction between real and artificial faces. Notably, in their study, although both stylized and human faces elicited an enhanced N170/VPP response between approximately 175–200 ms following stimulus onset, only human faces were associated with a sustained positive deflection emerging after 400 ms. It's important to highlight, however, the artificial face stimuli were highly stylized and low in perceptual realism (i.e., illustrated doll faces), thereby limiting the extent to which these findings can be generalized to contemporary hyperrealistic AI-generated faces.

The aim of the present study was to determine whether, despite observers' limited ability to behaviorally discriminate real faces from highly realistic AI-generated faces, neural signals would nonetheless reveal sensitivity to this distinction.

Indeed, the variability characterizing real human faces and GAN-generated faces arises from fundamentally different generative processes. While natural facial variability emerges from complex biological and developmental constraints, resulting in subtle asymmetries and highly idiosyncratic configurations, GAN-generated faces are shaped by algorithmic optimization over learned statistical regularities. As a consequence, synthetic faces often exhibit increased global symmetry and reduced extreme individuality compared to real faces, positioning them closer to the center of the learned face distribution. This structural regularity may enhance perceptual fluency and, in turn, elicit a heightened sense of familiarity, despite a reduction in authentic facial uniqueness. While a small number of neurophysiological studies have begun to explore neural responses to artificial or synthetic faces, very few ERP investigations have directly compared real faces with hyper-realistic (indistinguishable) AI-generated faces, and none, to our knowledge, have systematically examined the cortical generators of these potentials to clarify where, how, and at which stage of processing the brain may differentiate between the two stimulus classes. Furthermore, existing studies have not explicitly addressed whether neural differentiation between real and AI-generated faces varies as a function of face sex, leaving unresolved the role of biologically and socially salient facial attributes in this context.

In this study, we combined event-related potential (ERP) analyses with source reconstruction techniques to characterize the temporal and spatial dynamics underlying the neural processing of real and AI-generated faces. Specifically, we quantified the mean area amplitude of early occipito-temporal N170[15–17], intermediate frontal and posterior N250 and P300[18], and late anterior and posterior components (P/N400, LP;[19,20] across predefined time windows and electrode clusters. For each ERP component, repeated-measures ANOVAs were conducted to assess the effects of face type, face sex, electrode site, and hemisphere. To further elucidate the cortical generators associated with category-specific face processing, standardized and weighted low-resolution electromagnetic tomography (swLORETA[21] was applied across early, intermediate, and late temporal windows corresponding to perceptual, evaluative, and higher-order processing stages.

Materials and methods

Participants

Thirty participants (15 males, 15 females) aged between 20 and 32 years (*M* age = 21.7 years, *SD* = 2.5) took part in the study. The sample size was determined through a power analysis conducted with G*Power software for a repeated-measures ANOVA design (within-subject factors; $\alpha = 0.05$, power = 0.80, medium effect size). The present sample size was adequate to detect medium-to-large repeated-measures effects, while smaller higher-order interactions may require further investigation. All participants were right-handed (*M* laterality score = 0.73, SD = 0.17), as assessed by the Edinburgh handedness inventory for lateral preference (range − 1/+1), and reported normal or corrected-to-normal vision and normal hearing. Participants were recruited through *Sona System* and received academic credits for their participation. Exclusion criteria comprised speech or reading disorders, psychiatric, neurological, or neurodevelopmental conditions, history of epilepsy, substance abuse, or psychotropic medication use within 48 h before testing. All participants provided written informed consent before taking part in the experiment. The study was approved by the Ethics Committee of University of Milano-Bicocca (CRIP) as minimal-risk research on May 8 2024 (Protocol No. RM-2024-822). All procedures were conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. Four participants were discarded for excessive EEG artefacts (see later). The final sample had a mean age of 21.62 years (SD = 2.45) and a mean laterality score of 0.71 (SD = 0.18).

Stimuli

Stimuli consisted of 440 standardized, full-color facial images—220 real photographs and 220 StyleGAN2-generated faces—equally divided by gender (110 male, 110 female per category). All of them were showing a positive or neutral facial expression (see Fig. 1). Faces were chosen to match the participants' apparent age range (18–35 years) and were exclusively of Caucasian ethnicity to minimize potential confounding effects, particularly the Other-Race Effect. The real faces were sourced from the Flickr-Faces-HQ (FFHQ) dataset[5]. The selection criteria for real faces mirrored those applied to artificial faces, ensuring comparability in terms of gender, age range, and ethnicity. Faces included neck and the upper portion of chest. To maintain consistency between real and AI-generated stimuli, the images were balanced across several attributes: Facial symmetry (symmetrical vs. asymmetrical faces), Facial expressions (smiling vs. neutral), Hair characteristics (color, length), Facial accessories (make-up, glasses, earrings, and piercings), and Background types (see Table 1). Real and AI-generated faces were closely matched across key visual features, including symmetry, expression, hair color and length, accessories (e.g., makeup, glasses), and background type. This balance ensured that perceptual

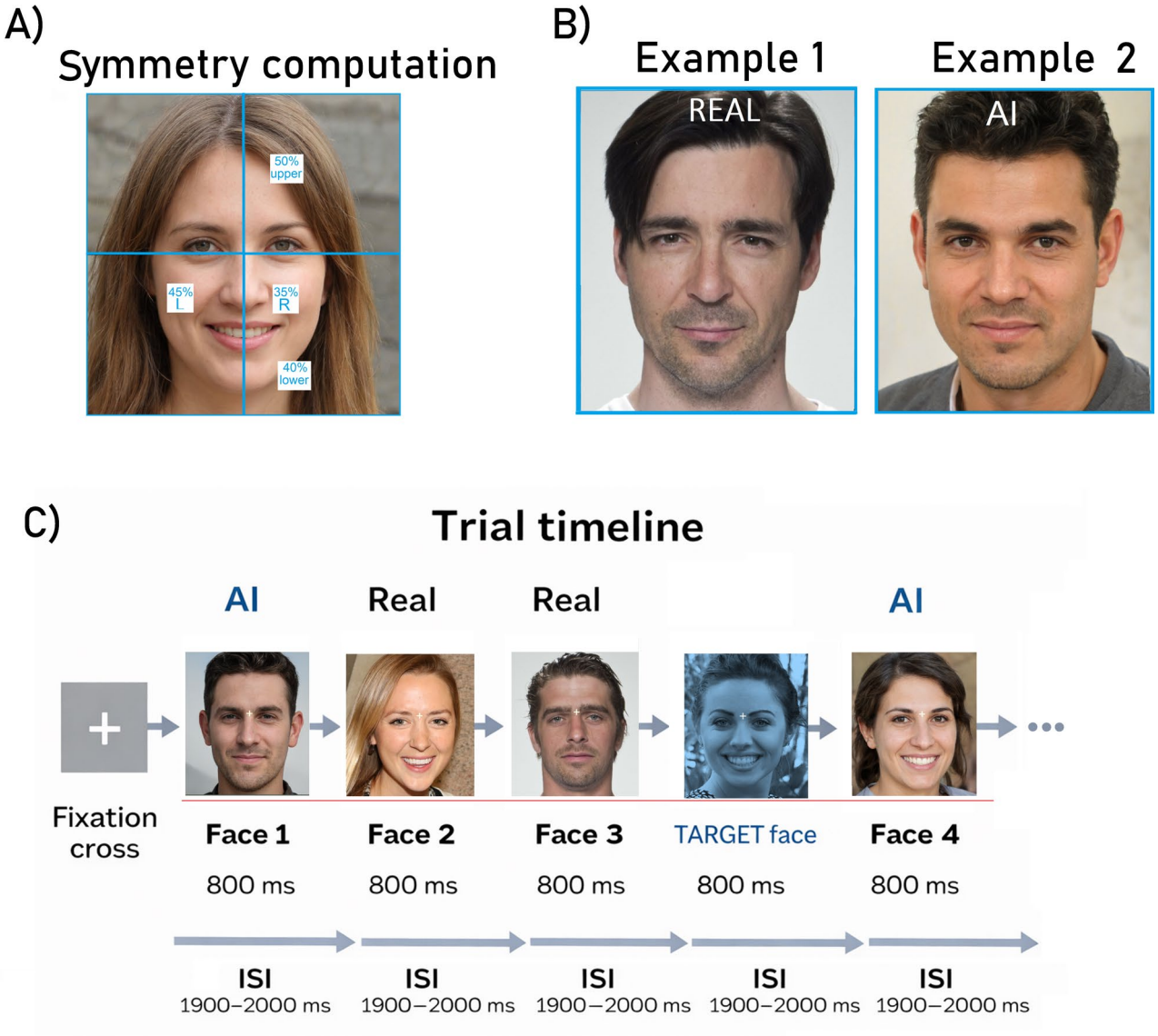

Fig. 1. (**A**) Face symmetry (position within framework, and orientation) was computed with respect to upper vs. lower and right vs. left space, and balanced across classes. The percentages represent the relative amount of visible facial oval contained within each quadrant of the frame. Here, "facial oval" refers only to the central facial region (forehead, cheeks, nose, mouth, and chin), excluding hair, ears, neck, and background. (**B**) Examples of male faces of the two classes. (**C**) Timeline of experimental procedure. The human faces were sourced from the open-access Flickr-Faces-HQ (FFHQ) dataset.

differences could be attributed to face origin (real vs. artificial), rather than confounding visual traits. Stimuli were isoluminant across classes, as shown by an ANOVA (p = .1, luminance range from 40 to 45 fL (full descriptive statistics are reported in the Supplementary Materials). Although no exact numerical equality was observed, the differences were minimal (~ 4 fL), suggesting that facial luminance was effectively matched, as closely as possible.

Stimulus validation

The stimuli were previously validated in a student sample drawn from the same cohort[22]. Fifty university students participated. Twenty-five participants (10 males, 15 females; mean age = 26.7 years) completed Study 1, in which they rated each face for aesthetic appeal and perceived familiarity on a 5-point Likert scale, without being informed that the faces could be either artificial or real. A separate group of 25 participants (8 males, 17 females; mean age = 24.9 years) completed Study 2, in which they classified each face as artificial or real (yes/no), with no information provided about the proportion of AI-generated faces. Both studies were administered via the SONA system and lasted approximately 20–25 min. Artificial faces were rated as more attractive than real faces (M = 3.12 vs. 2.31, SE = 0.04, $p < .0001$), and female faces were rated as more attractive than male faces (M = 3.02 vs. 2.41, SE = 0.04, $p < .0001$). Familiarity ratings showed the same pattern, with higher scores for artificial

Type	#	Symmetry			Smile		Hair color		
Female faces									
	Tot 440	Sym.	Asym. Right	Asym. Left	Open	Closed	Brown/ Dark	Blonde	Other
A.I.	110	58	26	26	70	40	66	42	2
Real	110	58	26	26	70	40	56	52	2
Male faces									
	Tot	Sym	Asym. Right	Asym. Left	Open	Closed	Brown/ Dark	Blonde	Mustaches
A.I.	110	58	30	22	57	53	97	13	0
Real	110	58	30	22	50	60	92	18	3

Type	Hair length		Paraphernalia		Hair length		Background	
Female faces								
	Long	Short/Tied	Makeup	Glasses	Long	Short/tied	Plain	Greenery
A.I.	87	23	74	4	87	23	44	23
Real	72	38	74	5	72	38	38	30
Male faces								
	Long	Short	Beard	Glasses	Long	Short	Plain	Greenery
A.I.	9	101	58	10	9	101	29	30
Real	10	100	58	10	10	100	40	24

Table 1. Balancing of perceptual features across AI-generated and real faces, as well as across female and male faces, to a lesser degree because of sexual dimorphism. Sym, symmetrical; Asym., asymmetry; Ear, earrings; Pier, Piercing. Symmetric faces were oriented strictly frontally, while slightly rotated faces, tilted slightly to the left or right, were classified as asymmetric.

than real faces (M = 2.70 vs. 2.24, SE = 0.02, $p < .001$) and for female than male faces (M = 2.54 vs. 2.41, SE = 0.02, $p < .001$). Participants failed to reliably identify AI-generated faces, performing well below chance (hit rate = 33%, SE = 0.01; chance = 50%), whereas classification of real faces slightly exceeded chance (hit rate = 62%, SE = 0.01), yielding a significant main effect of artificiality ($p < .0001$;[22]. A significant sex × artificiality interaction indicated particularly poor performance for AI-generated male faces relative to AI-generated female faces (hit rates = 29% vs. 37%, SE = 0.01). These findings suggest that artificial faces are not only poorly discriminated from real ones, but are systematically perceived as more familiar and aesthetically appealing. Participants' below-chance performance indicates a strong bias to attribute artificial faces to the category of "real," revealing an implicit trust in their realism. Crucially, this misattribution coexists with a devaluation of real faces, which are judged as less attractive and less familiar by comparison. Such a pattern implies that contemporary AI faces may exaggerate prototypical or fluency-enhancing features, leading to heightened aesthetic responses. Overall, the data point to a perceptual inversion whereby artificiality is normalized and realism becomes perceptually disadvantaged (Fig. 2).

Procedure

Eight stimulus sequences were programmed using *Gentask* (Compumedics Neuroscan, Charlotte, NC, USA). Each sequence comprised 60 stimuli, four of which were fictitious targets randomly distributed across the sequence. Each sequence included 14 faces per category (14 AI-generated male, 14 AI-generated female, 14 real male, 14 real female, totaling 56 faces per sequence), presented with equal probability and randomly intermixed. The last sequence was marginally shorter consisting of 48 faces (12 per category). Sequence order and response-hand assignment were counterbalanced across participants. Participants were instructed to keep their index finger resting on the response pad and to execute a button press whenever a target stimulus (a bluish-tinted face) was detected. To minimize habituation and order effects, sequence order was individually randomized. Furthermore, sequence randomization controlled for carry-over effects relative to target selection. Each stimulus was displayed for 800 ms, followed by an interstimulus interval (ISI) jittered between 1900 and 2000 ms (Fig. 1C). Stimuli subtended a visual angle of 6° × 6° (1024 × 1024 pixels; physical size: 12 × 12 cm) and were presented on a monitor positioned 114 cm from the participant. Participants were instructed to maintain central fixation throughout the task; accordingly, a small yellow fixation cross was continuously displayed at the center of the screen. Before EEG data acquisition, participants completed a structured training session to ensure task comprehension. The total duration of EEG recording was approximately 45 min.

EEG recordings

Brain bioelectric activity was recorded using a Compumedics quick cap neo-net with 128 sensors arranged according to the 10/5% system[23], with a sampling frequency of 512 Hz. Eye movements (saccades and blinks) were detected via vertical and horizontal electro-oculograms (VEOG and HEOG), via bipolar electrodes placed on the side of the eye (outer canthus), and below/above the right eye. Signals were referenced to a common average reference (CAR) to optimize source reconstruction. The EEG signals were filtered offline with a 0.01–70 Hz bandpass filter. Eye blinks were reduced using the covariance reduction method, and excessively noisy

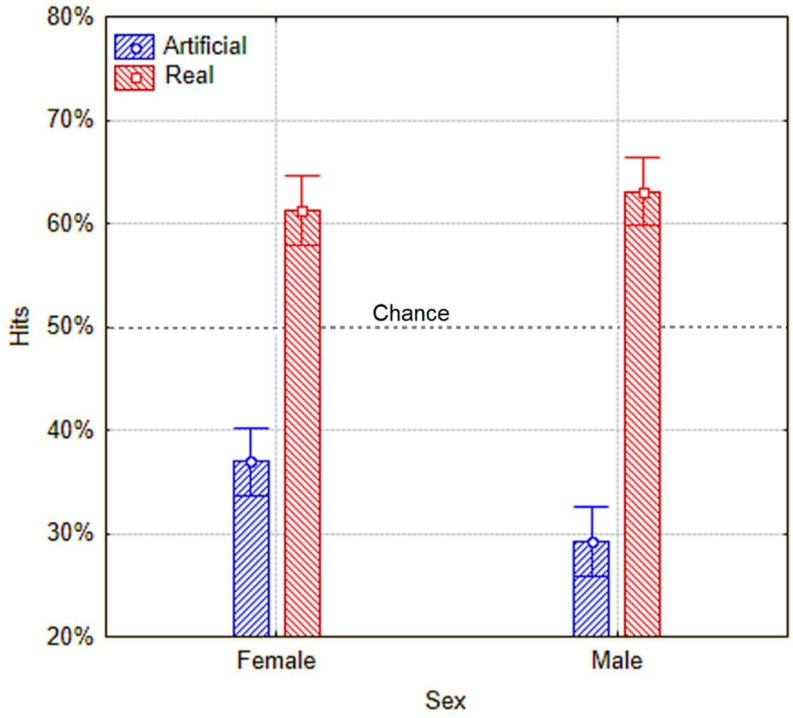

Fig. 2. Mean classification accuracy (%) for AI and real faces as a function of face sex. Error bars represent standard errors.

electrodes were interpolated with the four neighbouring electrodes. The EEG was filtered using the threshold method at ± 75 μV, and a 50 Hz notch filter was applied. The EEG epochs were synchronized with the onset of the stimuli and analysed using *Curry9* software (Compumedics Neuroscan), segmented into epochs extending from 100 ms pre-stimulus to 800 ms post-stimulus onset. Event-Related Potentials (ERPs) were extracted using the offline averaging technique. ERP data from fictitious rare targets were excluded from analysis. The electrophysiological data of four subjects were excluded after ERP averaging, because of excessive EEG/EOG artifacts and an insufficient number of trials (below 30%). The final sample comprised twenty-six participants (13 M, 13 F). The peak amplitude of occipito/temporal N1 was quantified in between 170 and 210 ms post-stimulus latency at PO7, PO8, PO9 PO10, P9, P10 electrode sites using *Eeprobe* software (ANT Neuro, Hengelo, The Netherlands). The mean area amplitude of anterior frontal N250 response was quantified in between 250 and 350 ms at AF3, AF4 site. In the same time window the mean area amplitude of posterior P300 was quantified at O1, O2, PO9, PO10, TP7, TP8, P3, P4 electrode sites. Anterior FN/P400 and posterior late positivity (LP) were quantified in between 350 and 450 ms at AF3, AF4, AFF1h, AFF2h, AFF3h, AFF4h, and O1, O2, PO9, PO10, TP7, TP8, P3, P4 electrode sites, respectively. The N170 component was quantified using peak amplitude due to its sharp and well-defined morphology, whereas later components were measured using mean amplitude within predefined time windows to better capture their broader temporal dynamics. For each ERP component mean area values underwent distinct repeated measures ANOVA, whose variability factors were: Face type (AI, Real), Face sex (Male, Female), Electrode (3/4 levels), Hemisphere (left, right). Tukey post-hoc comparisons were performed to assess differences among means. The effect-size for statistically significant factors was calculated with partial eta squared (η^2_p) and Greenhouse-Geisser epsilon correction was applied for ε values <1. Multiple ERP components were analyzed to capture distinct stages of face processing, consistent with standard practice. While the main effects of face type on N170 (p = .005), P300 (p = .0001), PN400 (p = .002), and LP (p = .00012) remained significant after conservative Bonferroni correction (α/5 = 0.01), supporting the robustness of the principal findings, multiple interactions should be interpreted with greater caution.

SwLORETA and source reconstruction

Standardized and weighted *Low-Resolution Electromagnetic Tomography* (swLORETA;[21,24], was applied to estimate intracortical generators explaining scalp-recorded potentials to different face classes in three different time windows corresponding to N170 component (170–210 ms), N250 and P300 response (250–350 ms) and LP and PN400 response (350–450 ms). 3D Isocolour topographical maps and SwLORETA were performed with *ASA 4.10.1* software (ANT Neuro, The Netherlands). LORETA is a linear distributed source localisation algorithm that provides discrete solutions to the EEG inverse problem by estimating the three-dimensional distribution of neuronal electrical activity with maximal similarity in orientation and strength across neighbouring neuronal populations, modelled as adjacent voxels. Source reconstruction was performed using the standardised weighted LORETA algorithm (swLORETA), which incorporates a singular value decomposition–based source field weighting procedure. The source space was defined using a grid spacing of 5 mm and an estimated signal-to-noise

ratio (SNR) of 3, which determines the degree of Tikhonov regularisation, with higher values corresponding to reduced spatial blurring. An SNR range of 3–4 has been shown to yield superior localisation accuracy across inverse problem evaluations. SwLORETA was applied to the grand-averaged group data to identify statistically significant electromagnetic dipoles ($p < .05$), with greater source magnitudes reflecting stronger activations. Data were automatically re-referenced to the mean reference as part of the LORETA procedure. A realistic boundary element model (BEM) was derived from a T1-weighted three-dimensional MRI dataset by segmenting brain tissue. The resulting BEM consisted of a homogeneous compartment comprising 3446 vertices and 6888 triangles. Source analysis was conducted using Advanced Source Analysis (ASA), which implements a three-layer realistic head model (scalp, skull, brain) generated with the BEM approach[25]. The model was defined by irregularly shaped boundaries approximated by triangulated meshes, with spatial resolution determined by the selected grid spacing. Segmentation was assumed to include current generators throughout the brain volume, encompassing both grey and white matter. Conductivity values for the scalp, skull, and brain were set to 0.33, 0.0042, and 0.33, respectively. Source reconstruction solutions provided by the Montreal Neurological Institute were projected onto the three-dimensional MRI of the Collins brain. Probabilities of source activation were computed for each independent EEG source using Fisher's F-test and reported on a unit scale in nanoamperes (nA), with larger values indicating higher statistical significance. Distinct colour scales were used to represent different strengths of the electromagnetic signals. Furthermore, source-reconstructed neural activity was analyzed using non-parametric Wilcoxon signed-rank tests (sign test equivalent), which were independently applied in the present study to perform pairwise comparisons across conditions.

Results

Electrophysiological data

Figure 3 shows the grand-average ERP waveforms recorded at anterior, temporo-parietal, and occipito-temporal electrode sites in response to faces from the four experimental categories. The face-sensitive N170 component exhibited its largest amplitude over the right occipito-temporal region. The posterior P300 and late positivity (LP), as well as the anterior N250 and PN400 components, were modulated by face category, despite all stimuli being presented as non-targets. The ANOVA performed on N170 peak amplitude values showed the significance of face type factor [$F(1, 25) = 10.28$, $p = .005$, $\eta^2_p = 0.29$, $\varepsilon = 1$], with greater amplitude in response to real ($M = -1.71$ µV, $SE = 2.68$) than AI faces ($M = -1.22$ µV, $SE = 2.49$) (Fig. 4). The ANOVA also yielded the significance of sex factor [$F(1, 25) = 7.12$, $p = .01$, $\eta^2_p = 0.22$, $\varepsilon = 1$] with larger responses to male ($M = -1.67$ µV, $SE = 2.60$) than female faces ($M = -1.26$ µV, $SE = 2.57$), and electrode factor [$F(2, 50) = 26.8$, $p = .00001$, $\eta^2_p = 0.52$, $\varepsilon = 1$] with larger responses over the occipito/temporal than lateral occipital sites. The ANOVA performed on N170 latency values showed the significance of face type factor [$F(1, 25) = 17.34$, $p = .0005$, $\eta^2_p = 0.41$, $\varepsilon = 1$], with faster latencies for AI ($M = 193$ ms, $SE = 0.001$) than real ($M = 194$ ms, $SE = 0.001$) faces. Given the 512 Hz sampling rate, the observed latency difference should be interpreted as a minimal but reliable shift within the temporal resolution of the recording system.

The ANOVA performed on anterior N250 response (250–350 ms) amplitudes showed the significance of face type x sex interaction [$F(1, 25) = 4.65$, $p = .04$, $\eta^2_p = 0.16$, $\varepsilon = 1$]. Post-hoc comparisons showed that for AI faces, mean N250 amplitudes were comparable in males ($M = -3.36$, $SE = 0.82$ µV) and females ($M = -3.38$ $SE = 0.86$ µV). In contrast, for real faces, males showed a reduced N250 amplitude ($M = -2.91$ $SE = 0.74$ µV) compared to females ($M = -3.29$ $SE = 0.75$ µV), as visible in maps of Fig. 5(left). Overall, these findings suggest a sex-dependent modulation of N250 responses to face realism, with a greater attenuation of N250 amplitude to real faces in males relative to females.

The ANOVA performed on posterior P300 response (250–350 ms) showed the significance of face type factor [$F(1, 25) = 22.24$, $p = .0001$, $\eta^2_p = 0.47$, $\varepsilon = 1$]. Mean P300 amplitudes (in µV) were higher in response to AI faces ($M = 5.55$, $SE = 1.92$) compared to real faces ($M = 4.91$, $SE = 1.85$). The ANOVA also yielded the significance of Electrode factor [$F(3, 75) = 58$, $p = .0001$, $\eta^2_p = 0.70$, $\varepsilon = 0.544$; adjusted p value = 0.0001]. Tukey post-hoc comparisons showed that the largest P300 responses were observed at occipital (O1–O2) and parieto-occipital (PO9–PO10) sites ($p = .0001$), with mean amplitudes of 8.78 µV ($SE = 2.37$ and 7.80 µV ($SE = 2.17$), respectively. Intermediate amplitudes were found at parietal electrodes (P3–P4) (3.03 $SE = 1.21$ µV), whereas the smallest responses were recorded at temporo-parietal sites (TP7–TP8) ($M = 1.30 \pm 0.80$ µV). Also significant was the factor hemisphere [$F(1, 25) = 7.25$, $p = .01$, $\eta^2_p = 0.22$, $\varepsilon = 1$], with larger mean amplitude over the right hemisphere ($M = 5.85$ $SE = 2.20$ µV) compared to the left hemisphere ($M = 4.60$ $SE = 1.97$ µV)(see Fig. 5 right). Mean P300 amplitudes (in µV) differed as a function of face type and sex [$F(1, 25) = 6.76$, $p = .01$, $\eta^2_p = 0.213$, $\varepsilon = 0.88$; adjusted p value = 0.015]. In response to AI faces, males showed a mean P300 amplitude of 5.62 µV ($SE = 1.41$), while females exhibited a comparable amplitude of 5.47 µV ($SE = 1.32$). In contrast, for real faces, P300 amplitudes were lower in males ($M = 4.77$ $SE = 1.38$ µV) and slightly higher in females (M = 5.05 $SE = 1.27$ µV).

The ANOVA performed on anterior frontal PN400 (350–450 ms) yielded the significance of face type factor [$F(1, 25) = 11.77$, $p = .002$, $\eta^2_p = 0.32$, $\varepsilon = 1$]. The mean NP400 amplitude was −2.95 µV (SE = 1.38) in the AI condition and −2.43 µV ($SE = 1.16$) in the Real condition, thus showing a larger anterior N400 to AI faces, as can be seen in waveforms of Fig. 3 and topographical maps of Fig. 6 (left). The ANOVA also yielded the significance of the face type x sex interaction [$F(1, 25) = 7.73$, $p = .01$, $\eta^2_p = 0.24$, $\varepsilon = 1$]. Post-hoc comparisons showed that all mean amplitudes were comparable across conditions and were larger than the NP400 to real male faces. The mean NP400 amplitude was −2.97 µV ($SE = 0.96$) for males and −2.94 µV ($SE = 1.02$) for females in the AI condition, and −2.16 µV ($SE = 0.78$) for males and −2.69 µV ($SE = 0.92$) for females in the Real condition. This effect can be appreciated in Fig. 6 (left).

The ANOVA performed on posterior LP (350–450 ms) showed the significance of face type factor [$F(1,25) = 20.68$, $p = .00012$, $\eta^2_p = 0.45$, $\varepsilon = 1$]. The late positivity (LP) showed a larger amplitude for AI faces

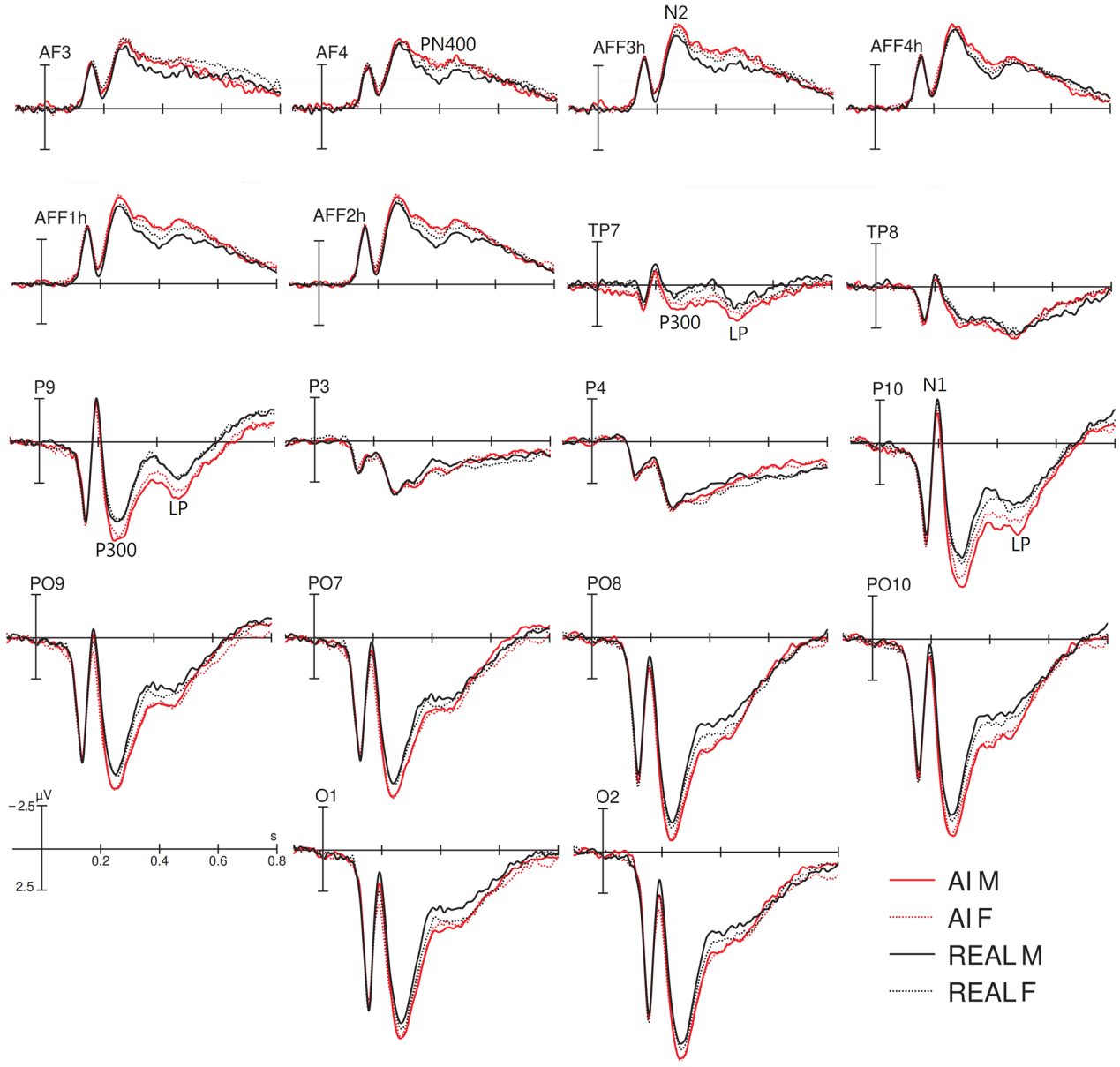

Fig. 3. Grand-average ERP waveforms recorded from anterior, anterior/central, temporo/parietal, parietal, occipito/temporal and occipital sites in response to AI-generated and real faces of both sexes.

($M = 3.52$ μV, $SE = 1.82$) compared to Real faces ($M = 2.89$ μV, $SE = 1.74$). A significant main effect of electrode was also observed [$F(3,75) = 17.11$, $p < .001$, $\eta^2_p = 0.41$, $\varepsilon = 0.47$, adjusted p value = 0.001]; post-hoc comparisons showed that LP was larger at occipital ($M = 4.68$ μV, SE = 1.94), and occipito/temporal sites ($M = 4.40$ μV, SE = 2.18), followed by parietal sites P3–P4 ($M = 2.34$ μV, $SE = 0.85$). The smallest LP amplitude was observed at temporo-parietal sites TP7–TP8 ($M = 1.41$ μV, $SE = 0.70$). See the waveforms of Fig. 3 and topographical maps in Fig. 6 (right). The ANOVA also yielded a significant hemisphere factor [$F(1,25) = 10.53$, $p = .003$, $\eta^2_p = 0.30$,$\varepsilon = 1$], with larger LP potentials over the right (3.90 μV, $SE = 1.86$) than left hemisphere (2.51 μV, SE = 2.05). Moreover, the interaction between face type and face sex was significant [$F(1,25) = 7.50$, $p = .011$, $\eta^2_p = 0.23$, $\varepsilon = 1$]: real female faces elicited smaller LP amplitudes than AI faces (< 0.001), but larger amplitudes than real male faces (< 0.001), with no difference between male and female AI faces, as can be appreciated form topographical maps depicted in Fig. 6 (right). For AI faces, the late positivity (LP) amplitude was 3.57 μV ($SE = 1.28$) for male faces and 3.48 μV ($SE = 1.32$) for female faces. For Real faces, the LP amplitude was 2.64 μV ($SE = 1.26$) for male faces and 3.14 μV ($SE = 1.24$) for female faces. Finally, a significant interaction between face type and electrode was found [$F(3,75) = 9.01$, $p < .001$, $\eta^2_p = 0.27$, $\varepsilon = 0.897$, adjusted p value = 0.0001] showing significant face type effect only at occipito/temporal sites (O1-2 and PO9-10) similarly to P300 component.

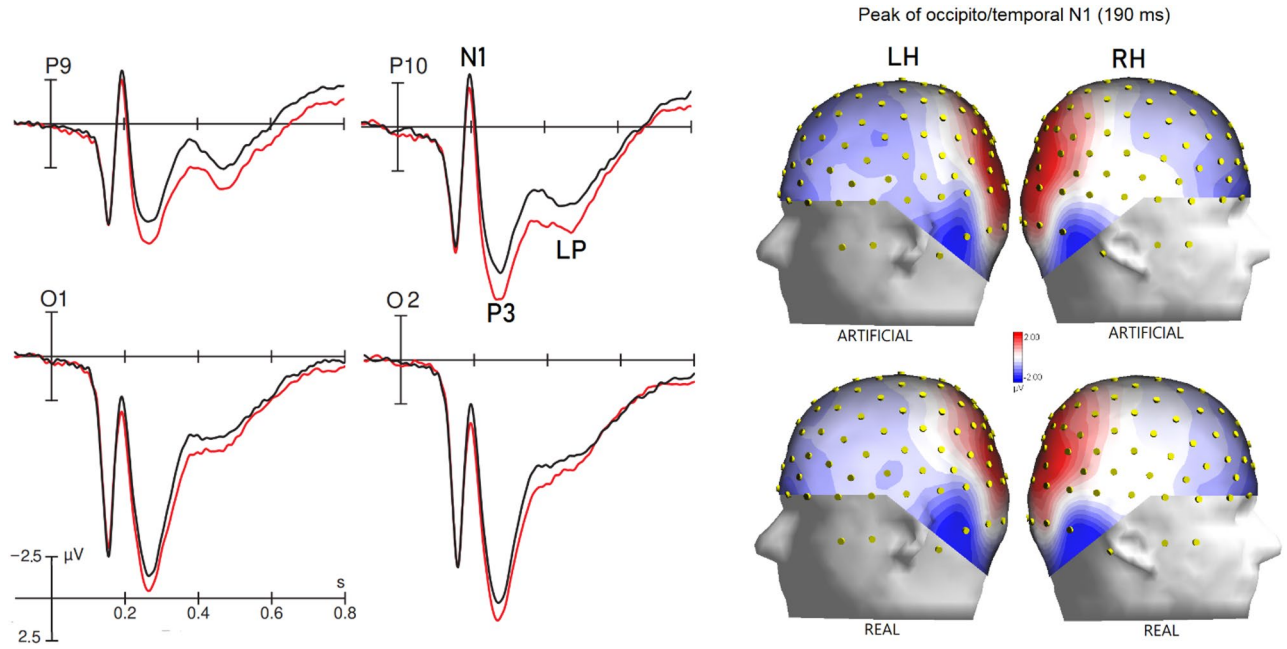

Fig. 4. (Left) Grand-average ERP waveforms recorded from left and right occipito/temporal and occipital sites in response to real (black line) and AI faces (red line. (Right) Isocolour scalp topographies depicting the distribution of N170 surface voltage across the left and right hemispheres in response to AI-generated versus real faces.

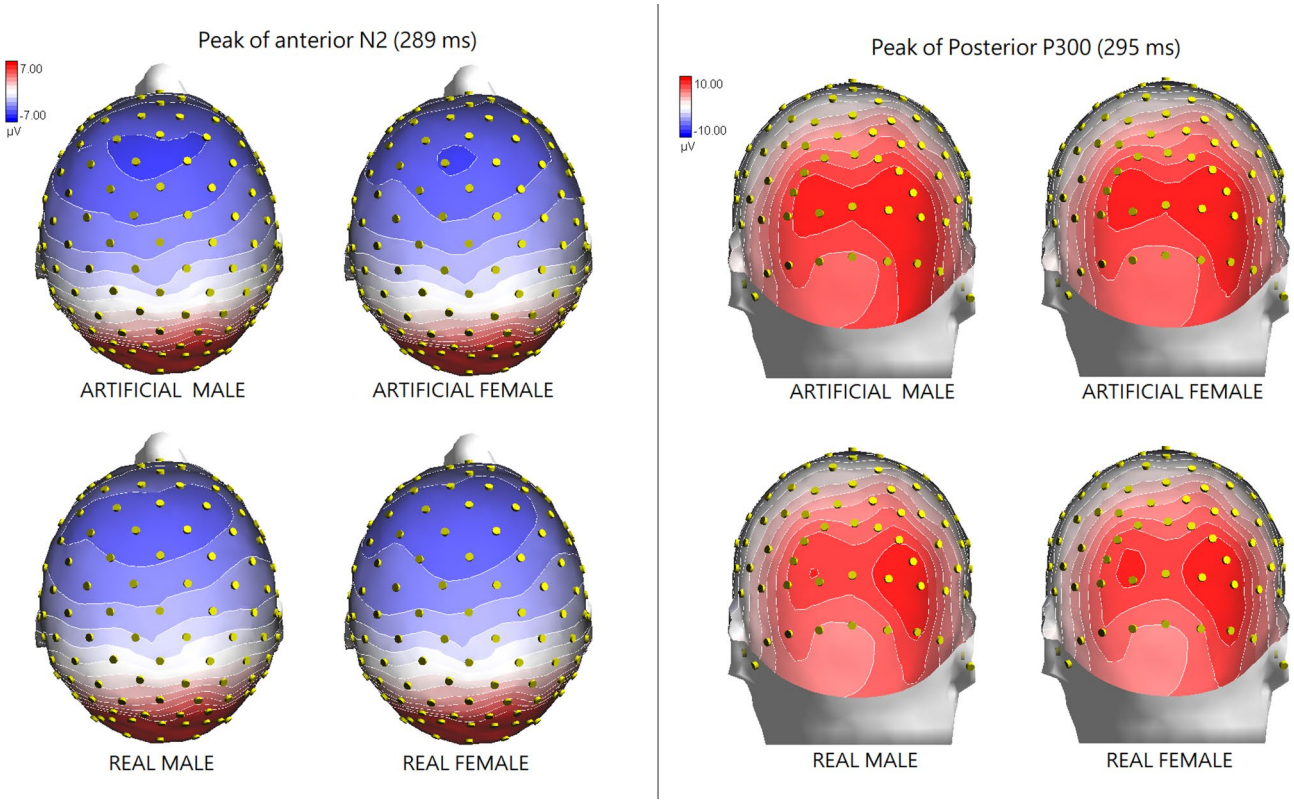

Fig. 5. (Left) Isocolour scalp topographies depicting the distribution of anterior N250 surface voltage at peak latency in response to faces of the four face types. The evaluative response showed larger potentials to AI faces. (Right) Isocolour scalp topographies depicting the distribution of posterior P300 surface voltage at peak latency in response to faces of the four face types. P300 was reduces for male than female real faces, while there was no sex effect for AI-generated faces.

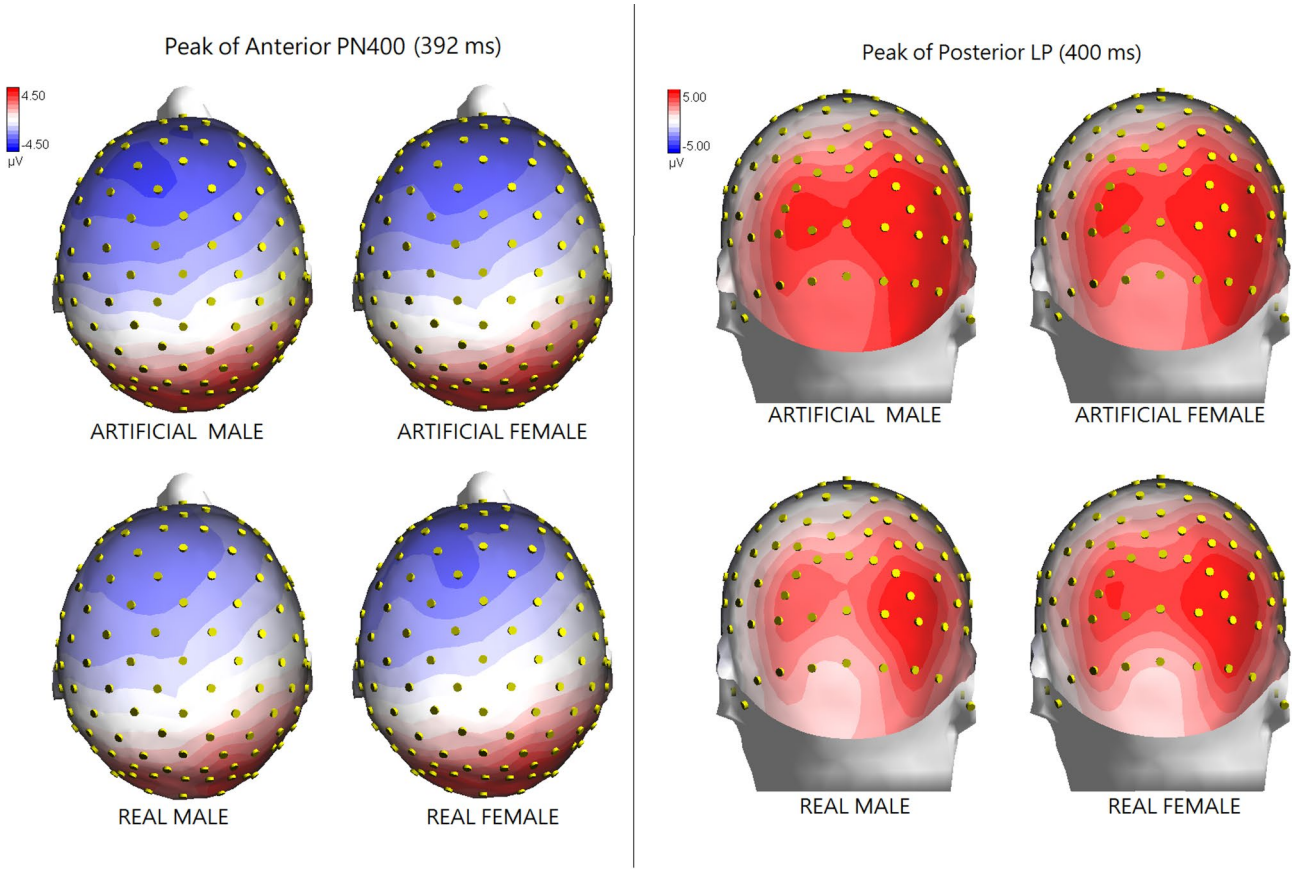

Fig. 6. (Left) Isocolour scalp topographies depicting the distribution of anterior N400 surface voltage at peak latency in response to faces of the four face types. (Right) Isocolour scalp topographies depicting the distribution of posterior LP surface voltage at peak latency in response to faces of the four face types. Both responses were significantly attenuated for real relative to AI faces, with no detectable sexual dimorphism in the latter condition.

swLORETA source reconstruction

Source reconstructed neural activity was analyzed using non-parametric Wilcoxon signed-rank tests (sign test equivalent) to perform pairwise comparisons across conditions. Results are reported for descriptive purposes rather than as a primary inferential framework, given the absence of individual-level variance estimates.

During the 170–210 ms time window, corresponding to the face-specific N170 response, neural activity within temporal cortical regions (fusiform gyrus, BA 19 and BA 21) appeared greater for real than AI faces (see Fig. 7, top), consistent with the involvement of occipital and fusiform face-processing areas[26,27]. This pattern qualitatively mirrored the enhanced amplitude of the occipitotemporal N170 component observed at the scalp level for real relative to AI faces (see Table 2, top, for a list of active electromagnetic dipoles). During the 250–350 ms time window, corresponding to the surface components N250 and P300, source localization of intra-cortical generators revealed differentiated patterns across stimulus categories (see Table 2; Fig. 7, middle). Artificial faces (artificial male, artificial female) and real female faces showed a broadly consistent distribution of activity within ventral temporal, posterior parietal, and limbic networks. In contrast, real male faces showed a partially distinct configuration of generators, characterized by the presence of a right insular source not observed in the other face categories. Across artificial male, artificial female, and real female faces, activity patterns included: robust engagement of ventral temporal regions, including the fusiform gyrus (BA20) and superior temporal gyrus (BA22/41); consistent involvement of limbic regions, particularly the parahippocampal gyrus (BA27/28) and posterior cingulate/cingulate cortex (BA31); and contributions from posterior parietal midline regions, notably the precuneus (BA19), commonly associated with the P3b complex. These generators showed comparable or relatively greater magnitudes across these three conditions, consistent with a stable P300/N250-related network configuration. Real male faces, by comparison, were associated with: the presence of a right insular source (BA13), not observed in the other conditions; relatively lower magnitudes of temporo-limbic generators, including fusiform and cingulate regions; and a comparatively reduced posterior parietal (precuneus) contribution. This source-level pattern qualitatively mirrored the attenuated posterior P300 and anterior N250 amplitudes observed in the ERP waveforms for the real male condition.

During the 350–450 ms time window, corresponding to the surface components PN400 and Late Positivity (LP), neural activity was predominantly distributed within limbic and paralimbic regions, including the uncus,

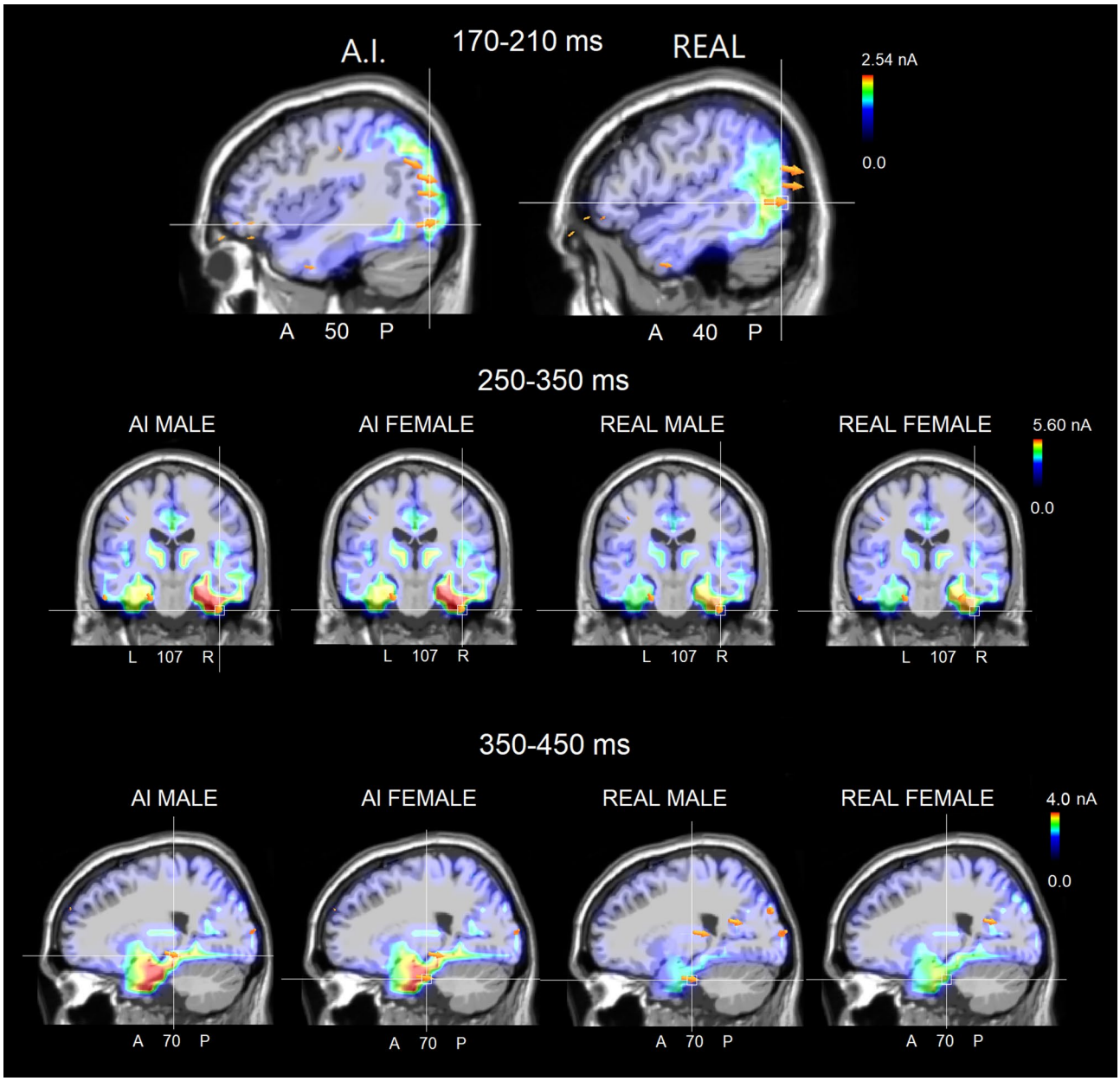

Fig. 7. (Top) Sagittal sections of swLORETA active within the N170 time window (approximately 170–210 ms) in response to AI vs. real faces. (Middle) Coronal sections of swLORETA active within the N250/P300 time window (approximately 250–350 ms) in response to AI vs. real faces of the two sexes. (Bottom) Sagittal sections of swLORETA active within the PN400/LP time window (approximately 350–450 ms) in response to AI vs. real faces of the two sexes. The electromagnetic dipoles appear as arrows and indicate the position, orientation and magnitude of the dipole modelling solution applied to the ERP waveform in the specific time window. L, left; R, right; numbers refer to the displayed brain slice in the MRI imaging plane.

parahippocampal gyrus, and cingulate cortex, as well as the insula, cuneus, supramarginal gyrus, and superior frontal cortex (Fig. 7, bottom). Across these regions, AI-generated faces showed relatively stronger and more spatially consistent activations, whereas real male faces were associated with comparatively attenuated neural responses, particularly within right limbic regions and frontal cortex. Real female faces generally showed intermediate response magnitudes, greater than those observed for real male faces but lower than those associated with AI-generated faces, especially within limbic regions. Overall, this source-level pattern suggests that the PN400/LP time window involved distributed affective–mnemonic networks and qualitatively mirrored the reduced neural responses observed for real male faces relative to AI-generated faces.

Overall, source-level statistical analyses were intended as complementary to sensor-level ERP findings. In the 170–190 ms window, Wilcoxon tests revealed reduced activation for AI compared to real faces in temporal regions (FDR-corrected, $p < .05$), possibly linked to visual N170 sources. Effect sizes ranged from medium to

Hem.	Lobe	Gyrus	BA	AI	Real
170–190 ms					
R	O	Middle occipital	19	2.54	2.41
R	T	Inferior temporal	19	< 0.5	2.39
L	O	Middle occipital	19	2.25	2.04
L	O	Precuneus	31	2.26	< 0.5
R	T	Middle temporal	21	1.19	1.25
L	F	Superior frontal	10	0.83	0.761
R	F	Superior frontal	10	0.83	0.739

Hem.	Lobe	Gyrus	BA	AI M.	AI F.	Real M.	Real F.
250–350 ms							
R	T	Inferior temporal	20	6.43	6.314	5.47	4.48
R	Limbic	Parahippocampal	27	6.16	6.05	5.19	5.43
R	Limbic	Cingulate	31	5.976	5.88	5.15	5.44
L	P	Precuneus	19	5.486	5.57	4.97	5.09
L	Limbic	Parahippocampal	28	5.367	5.39	4.62	4.51
L	T	Superior temporal	41	3.821	3.68	3.19	3.42
R	F	Medial frontal	10	2.185	1.84	2.04	1.85
350–450 ms							
R	Limbic	Uncus	20	4.44	4.33	2.81	3.55
R	Limbic	Parahippocampal	28	3.93	3.9	2.80	3.25
L	Limbic	Parahippocampal	28	3.59	3.67	3.51	2.63
R	Limbic	Cingulate	31	3.43	3.45	< 0.5	3.17
R	SubLobar	Insula	13	3.00	3.07	2.33	2.60
R	O	Cuneus	18	2.66	2.83	2.71	< 0.5
R	P	Supramarginal	40	2.55	2.52	2.04	2.41
L	F	Superior frontal	10	1.54	1.72	< 0.5	0.80

Table 2. Magnitude (in nA) of main electromagnetic dipoles found active in the three time windows considered. Hem, hemisphere; BA, Brodmann area; M., male; F., female.

large. In the 250–350 ms and 350–450 ms windows, real male faces showed a robust and sustained attenuation of neural activation relative to both AI-generated faces and real female faces, particularly within temporal, limbic, parietal, and frontal regions (all $p < .05$, FDR-corrected), indicating a selective and temporally stable reduction associated with real male faces.

Discussion

Stimulus validation indicated a marked asymmetry in face authenticity judgments: while real human faces were identified as such at slightly above-chance levels, AI-generated faces systematically failed to be recognized as artificial. This pattern suggests a default bias toward attributing humanness to facial stimuli, whereby observers were more efficient at confirming human authenticity than at detecting artificial origin. According to the literature, hyperrealistic synthetic faces appear to exploit perceptual heuristics underlying face processing—such as reliance on prototypical structure and global coherence, leading artificial faces to be misclassified as real[9,10]. This limitation reflects fundamental constraints in human face expertise, with realism judgments guided more by perceived plausibility than by sensitivity to generative artifacts, even when observers are explicitly instructed to perform authenticity decisions[13].

In contrast, overall, the face-specific N170 response was slightly larger and slower for real than for AI-generated faces, despite the fact that AI faces were absolutely not recognized as AI-made at conscious and behavioral level. This visual evoked perceptual component showed as expected sensitivity to face category[19,28,29], indicating differential perceptual demands across conditions. Furthermore, N170 was found to be slightly larger to male than female faces. In particular, real male faces potentially perceived as less aesthetically pleasing or less prototypical according to stimulus validation[22] may have required greater early perceptual resources, consistent with accounts linking increased N170 amplitude to enhanced visual processing demands for less optimal or unfamiliar faces[30,31]. Quite similarly Trujillo et al.,[32] found that the posterior N170 (150–225 ms) face-evoked ERP component was smaller in response to high-attractive and averaged faces than to low-attractive faces. In contrast, previous studies showed that robotic vs. human faces did not seem to modulate N170 and early stages of face information processing[33,34]. In our study, face-specific N170 peaked at approximately 190 ms. This latency is slightly later than the ~ 170 ms typically reported in the literature when faces serve as task-relevant targets[19,26], likely because in the present paradigm faces were not attended (were indeed non-target stimuli), thereby favoring a more automatic mode of encoding.

At anterior frontal sites, the anterior N250 and PN400 showed convergent sensitivity to face realism and sex, supporting their interpretation as markers of early and late evaluative and aesthetic appraisal processes.

Both components were enhanced for AI faces and attenuated for real male faces, suggesting reduced aesthetic appreciation for the latter. The absence of sex differences for AI faces further indicates that aesthetic evaluation may be overridden when faces are perceived as artificial, or that sexual dimorphism is reduced for AI male faces, or they are slightly effeminate. At later stages, the posterior P300 and LP exhibited highly similar modulation patterns, consistent with their association with stimulus perceived familiarity. Both components were enhanced for AI faces and showed maximal responses over occipito-temporal regions, with additional sex-related modulations restricted to real faces. Notably, real female faces elicited larger LP amplitudes than real male faces, suggesting greater sense of familiarity or beauty. Taken together, these results indicate that increased familiarity, indexed by enhanced P300 and LP responses, tends to co-occur with greater aesthetic appreciation, reflected in larger anterior N250 and PN400 amplitudes. Conversely, stimuli perceived as less familiar or less aesthetically salient—such as real male faces—elicited reduced evaluative responses and increased early perceptual processing demands, pointing to a dynamic interaction between familiarity and aesthetic appraisal across multiple stages of face processing. From a different perspective, it is noteworthy that Pérez-Arenas et al.,[35] reported larger P300 amplitudes elicited by robotic faces compared to human faces, an effect they interpreted as reflecting enhanced attentional allocation and increased cognitive processing demands for artificial social stimuli relative to biologically relevant ones. Overall, the data showed slightly larger N170 responses to real than AI faces, especially male faces.

N250 and aesthetic appraisal of faces

The present results show that the anterior N250 amplitude was modulated by face realism as a function of sex, with comparable responses to artificial faces in males and females, but a relative attenuation of the N250 to real faces in males. This pattern suggests that early face-sensitive neural responses are differentially engaged depending on both stimulus characteristics and observer sex. Within this framework, the modulation of the anterior N250 can be interpreted in relation to early evaluative and affective processes, rather than purely structural face encoding. Previous ERP studies have shown that anterior N2/N250 components are sensitive to the motivational and affective significance of conspecifics, including faces perceived as socially relevant, emotionally salient, or aesthetically appealing. For instance, enhanced anterior N2/N250 responses have been reported for liked, trustworthy, or attractive faces, supporting the idea that this component indexes early affective appraisal of socially meaningful stimuli[36,37].

Importantly, the N250 has also been linked to the processing of "lovable" or positively valued conspecifics[38], reflecting rapid access to affective representations associated with faces that elicit approach-related responses. This interpretation is supported by ERP work showing that the anterior N250 is selectively enhanced by highly "lovable" conspecific cues, most notably infant faces and baby-schema signals, which automatically attract attention and engage early evaluative mechanisms. Proverbio and colleagues repeatedly reported stronger anterior N250 responses to infant than adult faces, consistent with an innate affiliative bias toward biologically salient social stimuli[17]. Such findings suggest that the N250 does not merely reflect identity-related processing, but also integrates affective and motivational dimensions of face perception.

The present ERP results converge with the behavioral validation data showing that artificial faces and female faces[37] were perceived as more attractive, supporting the interpretation that enhanced or preserved N250 responses may reflect greater aesthetic appreciation. The attenuation of the N250 to real faces in males may therefore indicate reduced affective engagement with stimuli judged as less attractive, consistent with models positing that early face-related ERPs are shaped by subjective value and social relevance. As for AI faces' valence, a recent study using GAN-generated faces found that artificial faces were indistinguishable from real ones and were even judged as more trustworthy than real faces[10]. Moreover, another study showed that artificial faces of White individuals were judged as real more often than faces of actual real people[8], phenomenon termed *AI hyperrealism*.

Source reconstruction of the P300/N250 time window (250–350 ms) revealed a striking dissociation in the neural processing of real male faces compared with all other categories. Artificial male, artificial female, and real female faces consistently engaged a canonical ventral–limbic face-evaluation network, including fusiform and superior temporal cortices, the parahippocampal gyrus, posterior cingulate (BA31), and the precuneus—generators that collectively support robust P3b and anterior N250 responses. In contrast, real male faces uniquely recruited the right insula (BA13), a region absent in all other conditions and commonly associated with interoceptive and affective monitoring, bodily discomfort, and the detection of atypical or aversive stimuli[39–42]. This insular shift was accompanied by reduced magnitudes in fusiform, limbic, and parietal generators, paralleling the markedly attenuated P300 and N250 observed in the sensor-level data. Together, these findings suggest that real male faces were processed as less prototypical (more distinctive) or aesthetically coherent, triggering an insula-mediated evaluative pathway rather than the standard ventral temporal–posterior parietal network engaged by artificial and female faces. Importantly, validation data revealed a systematic shift in inter-eye distance (IED) in AI-generated faces, with larger IED in males and smaller IED in females relative to real faces, resulting in a compression of typical sex-related configurational differences (see the Supplementary file). This pattern suggests that GAN-based averaging may reduce natural facial dimorphism, increasing overall prototypicality. Within this framework, real male faces—preserving more natural configurational variability—may be processed as less prototypical, contributing to their distinct neural signature. Second, we acknowledge that this effect may also be influenced by specificity of current GAN-based face generation. In particular, generative models may differentially capture the statistical regularities of male versus female faces, possibly due to biases in training datasets or differences in within-category variability. As a result, GAN averaging procedures may produce representations that are more prototypical for female than male faces, leading to asymmetric neural responses when compared to real stimuli.

Conclusions

Extensive evidence shows that facial averageness and perceived familiarity reliably increase attractiveness, consistent with both the mere-exposure effect[43] and prototypical face processing[44]. Averageness and symmetry are good candidates for biologically based standards of beauty[45]. In their seminal study Langlois and Roggman[44] demonstrated that mathematically averaged composite faces—generated from multiple male and female faces—were consistently judged as more attractive than most of the individual faces contributing to the composites. Moreover, attractiveness increased monotonically as more faces were averaged, indicating that facial attractiveness is closely linked to typicality and central tendency within the face category. Vokey and Read[46] showed that facial familiarity covaries positively with attractiveness and likability, indicating that more typical/familiar faces tend to be perceived as more attractive, even though they are less well discriminated in recognition tasks.

Therefore, it is possible that AI-generated faces are perceived as more attractive because they appear more familiar, reflecting algorithmic averaging processes that enhance prototypicality, as opposed to genetically driven mechanisms that promote individual uniqueness. Given that all stimuli were technically unfamiliar to the participants, the perceived familiarity associated with AI-generated faces likely stems from specific 'AI-driven' structural properties. This suggests that the synthetic nature of these faces (their 'AI-ness') may inherently produce a sense of familiarity, possibly due to the optimization of prototypical features that the brain recognizes as more common or attractive. Despite evidence that AI-generated faces have surpassed the uncanny valley and are judged to be indistinguishable from, and even more trustworthy than, real faces[10], neural processing remains attuned to the difference (see also[7,11,13,14] revealing markers consistent with an increased perception of familiarity. Future work should assess whether these effects persist or are even enhanced under explicit authenticity-detection tasks.

Study limitations

It should be noted that behavioral and EEG data were collected from independent samples to maintain participant naivety regarding the stimuli; while these samples were drawn from the same demographic and academic cohort, this design precludes a direct within-subject correlation between individual behavioral performance and neural activity. Furthermore, while we used exclusively Caucasian faces with positive or neutral expressions, the extent to which these findings generalize to other ethnicities or affective conditions remains to be established.

Data availability

The datasets generated and/or analyzed during the current study are publicly available in the Bicocca Open Archive Research Data repository (Proverbio & Dosaikina, 2026), Version 1, [https://doi.org/10.17632/kjv4dckx8m.1](https:/doi.org/10.17632/kjv4dckx8m.1). To access the archive, please contact the corresponding author.

Received: 15 January 2026; Accepted: 22 June 2026
Published online: 30 June 2026

References

1. Dawel, A., Miller, E. J., Horsburgh, A. & Ford, P. A systematic survey of face stimuli used in psychological research 2000–2020. *Behav. Res. Methods*. **54** (4), 1889–1901. https://doi.org/10.3758/s13428-021-01705-3 (2022).
2. Crookes, K. et al. How well do computer-generated faces tap face expertise? *PLoS One*. **10** (11), e0141353. https://doi.org/10.1371/journal.pone.0141353 (2015).
3. Schindler, S., Zell, E., Botsch, M. & Kissler, J. Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Sci. Rep.* **7**, 45003. https://doi.org/10.1038/srep45003 (2017).
4. Balas, B. & Pacella, J. Artificial faces are harder to remember. *Comput. Hum. Behav.* **52**, 331–337. https://doi.org/10.1016/j.chb.2015.06.018 (2015).
5. Karras, T., Laine, S. & Aila, T. A style-based generator architecture for generative adversarial networks. *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*. **4401**, 4410 (2019).
6. Goodfellow, I. et al. Generative adversarial networks. *Commun. ACM*. **63** (11), 139–144. https://doi.org/10.1145/3422622 (2020).
7. Moshel, M. L., Robinson, A. K., Carlson, T. A. & Grootswagers, T. Are you for real? Decoding realistic AI-generated faces from neural activity. *Vis. Res.* **199**, 108079. https://doi.org/10.1016/j.visres.2022.108079 (2022).
8. Miller, E. J. et al. AI hyperrealism: Why AI faces are perceived as more real than human ones. *Psychol. Sci.* **34** (12), 1390–1403 (2023).
9. Tucciarelli, R., Vehar, N., Chandaria, S. & Tsakiris, M. On the realness of people who do not exist: The social processing of artificial faces. *iScience* **25** (12), 105441. https://doi.org/10.1016/j.isci.2022.105441 (2022).
10. Nightingale, S. J. & Farid, H. AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proc. Natl. Acad. Sci. U S A*. **119** (8), e2120481119. https://doi.org/10.1073/pnas.2120481119 (2022).
11. Chen, Y., Stephani, T., Bagdasarian, M.T. et al. Realness of face images can be decoded from non-linear modulation of EEG responses. *Sci Rep* **14**, 5683. https://doi.org/10.1038/s41598-024-56130-1. (2024).
12. Fitzpatrick, S. et al. Neural correlates of the uncanny valley effect for robots and hyper-realistic masks. *PLoS One*. **20**(3), e0311714. https://doi.org/10.1371/journal.pone.0311714 (2025).
13. Tarchi, P., Lanini, M. C., Frassineti, L. & Lanatà, A. Real and deepfake face recognition: An EEG study on cognitive and emotive implications. *Brain Sci.* **13** (9), 1233. https://doi.org/10.3390/brainsci13091233 (2023).
14. Wheatley, T., Weinberg, A., Looser, C., Moran, T. & Hajcak, G. Mind perception: Real but not artificial faces sustain neural activity beyond the N170/VPP. *PloS One*. **6** (3), e17960. https://doi.org/10.1371/journal.pone.0017960 (2011).
15. Bentin, S., Golland, Y., Flevaris, A., Robertson, L. C. & Moscovitch, M. Processing the trees and the forest during initial stages of face perception: electrophysiological evidence. *J. Cogn. Neurosci.* **18** (8), 1406–1421 (2006).
16. Jacques, C. et al. The inferior occipital gyrus is a major cortical source of the face-evoked N170: Evidence from simultaneous scalp and intracerebral human recordings. *Hum. Brain Mapp.* **40** (5), 1403–1418. https://doi.org/10.1002/hbm.24455 (2019).
17. Proverbio, A. M., Riva, F., Zani, A. & Martin, E. Is it a baby? Perceived age affects brain processing of faces differently in women and men. *J. Cogn. Neurosci.* **23**(11), 3197–3208 https://doi.org/10.1162/jocn_a_00041 (2011).
18. Nelson, C. M., Xia, M. & Hudac, C. M. Influence of daily experiences on ERP correlates of face perception of ones' romantic partner. *Soc. Neurosci.* **20** (4), 219–228. https://doi.org/10.1080/17470919.2025.2595019 (2025).

19. Proverbio, A. M., Riva, F., Martin, E. & Zani, A. Face coding is bilateral in the female brain. *PLoS One*. **5** (6), e11242. https://doi.org/10.1371/journal.pone.0011242 (2010).
20. Yovel, G. & Paller, K. A. The neural basis of the butcher-on-the-bus phenomenon: when a face seems familiar but is not remembered. *Neuroimage* **21** (2), 789–800. https://doi.org/10.1016/j.neuroimage.2003.09.034 (2004).
21. Palmero-Soler, E., Dolan, K., Hadamschek, V. & Tass, P. A. swLORETA: a novel approach to robust source localization and synchronization tomography. *Phys. Med. Biol./Phys. Med. Biol.* **52**(7), 1783–1800. https://doi.org/10.1088/0031-9155/52/7/002 (2007).
22. Shabani, N. & Proverbio, A. M. Deceptive Realism: How AI-Generated Faces Shape Human Social Perception. NeuroMi Conference in Artificial Intelligence for Neuroscience: from basic research to clinical practice, Milan, October 15th (2025).
23. Oostenveld, R. & Praamstra, P. The 5% electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* **112** (4), 713–719. https://doi.org/10.1016/s1388-2457(00)00527-7 (2001).
24. Pascual-Marqui, R. D., Michel, C. M. & Lehmann, D. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophysiol.* **18** (1), 49–65. https://doi.org/10.1016/0167-8760(84)90014-x (1994).
25. Zanow, F. & Knösche, T. R. ASA–Advanced source analysis of continuous and event-related EEG/MEG signals. *Brain Topogr.* **16** (4), 287–290 (2004).
26. Rossion, B., Joyce, C. A., Cottrell, G. W. & Tarr, M. J. Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage*. **20**(3), 1609–1624. https://doi.org/10.1016/j.neuroimage.2003.07.010 (2003).
27. Itier RJ, Taylor MJ. Source analysis of the N170 to faces and objects. *Neuroreport.* **15**(8), 1261-5 (2004).
28. Dalle Nogare, L. & Proverbio, A. M. Emojis vs. facial expressions: An electrical neuroimaging study on perceptual recognition. *Soc. Neurosci.* **18** (1), 46–64. https://doi.org/10.1080/17470919.2023.2203949 (2023).
29. Reinke, P., Deneke, L. & Ocklenburg, S. Asymmetries in event-related potentials part 1: A systematic review of face processing studies. *Int. J. Psychophysiol.* **202**, 112386. https://doi.org/10.1016/j.ijpsycho.2024.112386 (2024).
30. Sadeh, B. & Yovel, G. Why is the N170 enhanced for inverted faces? An ERP competition experiment. *Neuroimage* **53** (2), 782–789. https://doi.org/10.1016/j.neuroimage.2010.06.029 (2010).
31. Vizioli, L., Foreman, K., Rousselet, G. A. & Caldara, R. Inverting faces elicits sensitivity to race on the N170 component: a cross-cultural study. *J. Vis.* **10** (1), 151–123 (2010).
32. Trujillo, L. T., Jankowitsch, J. M. & Langlois, J. H. Beauty is in the ease of the beholding: a neurophysiological test of the averageness theory of facial attractiveness. *Cogn. Affect. Behav. Neurosci.* **14** (3), 1061–1076. https://doi.org/10.3758/s13415-013-0230-2 (2014).
33. Abubshait, A., Weis, P. P., Momen, A. & Wiese, E. Perceptual discrimination in the face perception of robots is attenuated compared to humans. *Sci. Rep.* **13**, 21419 (2023).
34. Geiger, A. R. & Balas, B. Robot faces elicit responses intermediate to human faces and objects at face-sensitive ERP components. *Sci. Rep.* **11**, 24095 (2021).
35. Pérez-Arenas, X., Rivera-Rei, Á. A., Huepe, D. & Soto, V. Do we view robots as we do ourselves? Examining robotic face processing using EEG. *Brain Sci.* **16** (1), 9. https://doi.org/10.3390/brainsci16010009 (2026).
36. Schacht, A. & Sommer, W. Emotions in word and face processing: early and late cortical responses. *Brain Cogn.* **69** (3), 538–550. https://doi.org/10.1016/j.bandc.2008.11.005 (2009).
37. Zhang, Y. et al. Identifying cognitive preferences for attractive female faces: an event-related potential experiment using a study-test paradigm. *J. Neurosci. Res.* **89** (11), 1887–1893. https://doi.org/10.1002/jnr.22724 (2011).
38. Proverbio, A. M., Adorni, R., Zani, A. & Trestianu, L. Sex differences in the brain response to affective scenes with or without humans. *Neuropsychologia* **47** (12), 2374–2388. https://doi.org/10.1016/j.neuropsychologia.2008.10.030 (2009).
39. Terasawa, Y., Shibata, M., Moriguchi, Y. & Umeda, S. Anterior insular cortex mediates bodily sensibility and social anxiety. *Soc. Cogn. Affect. Neurosci.* **8** (3), 259–266. https://doi.org/10.1093/scan/nss108 (2013).
40. Uddin, L. Q., Nomi, J. S., Hébert-Seropian, B., Ghaziri, J. & Boucher, O. Structure and function of the human insula. *J. Clin. Neurophysiol.* **34** (4), 300–306. https://doi.org/10.1097/WNP.0000000000000377 (2017).
41. Garfinkel, S. N. & Critchley, H. D. Interoception, emotion and brain: new insights link internal physiology to social behaviour. Commentary on:: Anterior insular cortex mediates bodily sensibility and social anxiety by Terasawa et al. (2012). *Soc. Cogn. Affect. Neurosci.* **8** (3), 231–234. https://doi.org/10.1093/scan/nss140 (2013).
42. Zhang, R., Deng, H. & Xiao, X. The Insular Cortex: An Interface Between Sensation, Emotion and Cognition. *Neurosci. Bull.* **40** (11), 1763–1773. https://doi.org/10.1007/s12264-024-01211-4 (2024).
43. Zajonc, R. B. Attitudinal effects of mere exposure. *J. Personal. Soc. Psychol.* **9** (2), 1–27. https://doi.org/10.1037/h0025848 (1968).
44. Langlois, J. H. & Roggman, L. A. Attractive faces are only average. *Psychol. Sci.* **1** (2), 115–121. https://doi.org/10.1111/j.1467-9280.1990.tb00079.x (1990).
45. Rhodes, G. The evolutionary psychology of facial beauty. *Annu. Rev. Psychol.* **57**, 199–226. https://doi.org/10.1146/annurev.psych.57.102904.190208 (2006).
46. Vokey, J. R. & Read, J. D. Familiarity, memorability, and the effect of typicality on the recognition of faces. *Mem. Cognit.* **20** (3), 291–302. https://doi.org/10.3758/bf03199666 (1992).

Acknowledgements

The authors gratefully acknowledge Giulia Giani for creating and validating the stimuli, Nafiseh Shabani for her assistance with stimulus validation and data acquisition, and Giulia Gnecchi and Yldjana Dishi for their support with experimental programming.

Author contributions

A.M.P.: Conceptualization, writing-review and editing, writing-original draft, funding, project administration, formal analysis, data curation, statistical analyses, visualization. M.D.: Visualization, formal analysis, data curation, statistical analyses, writing-original draft. Both authors contributed to data interpretation.

Funding

This research was supported by 56821-2024-ATE-0036 grant from University of Milano-Bicocca entitled Neural processing of artificially generated (A.I.) social entities. We also gratefully acknowledge the support provided by the Italian Ministry of University and Research under Grant No. 2023-NAZ-0206, PsyFuture—Dipartimento di Eccellenza 2023–2027, awarded to the Department of Psychology of the University of Milano-Bicocca.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-59487-7.

Correspondence and requests for materials should be addressed to A.M.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.