Cross-Subject EEG Feedback for Implicit Image Generation

Carlos de la Torre-Ortiz, Michiel M. Spapé, Niklas Ravaja, and Tuukka Ruotsalo

***Abstract*—Generative models are powerful tools for producing novel information by learning from example data. However, the current approaches require explicit manual input to steer generative models to match human goals. Furthermore, how these models would integrate implicit, diverse feedback and goals of multiple users remains largely unexplored. Here, we present a first-of-its-kind system that produces novel images of faces by inferring human goals directly from cross-subject brain signals while study subjects are looking at example images. We report on an experiment where brain responses to images of faces were recorded using electroencephalography in 30 subjects, focusing on specific salient visual features (VFs). Preferences toward VFs were decoded from subjects' brain responses and used as implicit feedback for a generative adversarial network (GAN), which generated new images of faces. The results from a follow-up user study evaluating the presence of the target salient VFs show that the images generated from brain feedback represent the goal of the study subjects and are comparable to images generated with manual feedback. The methodology provides a stepping stone toward humans-in-the-loop image generation.**

***Index Terms*—Brain–computer interfaces, electroencephalography (EEG), generative models, image generation.**

I. Introduction

GENERATIVE image models have recently enabled creative tasks by exhibiting the capability to produce previously nonexisting visual information. However, user control over specific visual features (VFs) remains challenging. For example, while these models can generate photorealistic human faces, users may find it challenging to manipulate subjective semantic features, such as perceived gender. Many approaches have tackled the issue of controlling generative models, relying on manually provided input and additional model training using separately labeled data [56], [70], [75]. Others have proposed interface designs that allow the expression of goals via text [49], sketching [17], or other example images [13]. However, it is unrealistic to assume an interface design that allows for manipulating each possible salient feature in an image. Alternatives like text-to-image models offer broader control but provide very coarse control over features. In addition, they often depend on the user's ability to describe their intended output accurately or require prompt engineering [35]. Even if we could train generative models where we could adjust every VF, we would face challenges when considering the subjective nature of visual perception: what we perceive as "old" or "young" changes as we age, and more abstract concepts (e.g., "trustworthy") can vary widely. Therefore, gathering feedback from multiple individuals is essential to represent a concept as conceived and judged by a group of users.

In all cases, naturally expressing whether an image displays the VFs we associate with a semantic label is significantly easier than articulating reasons for our judgment. Human-in-the-loop generative systems iteratively leverage this capability, constantly refining the model's output to align with the desired outcome. However, manually providing such iterative feedback can be laborious and impractical, particularly for systems that require dozens of iterations to converge to even simple VFs [68]. What if we could harness the swiftness of our instinctive judgments to infer the presence of desired VFs? Visual cognition allows us to evaluate stimuli in fractions of a second [34], indicating an opportunity to leverage these immediate responses or "gut feelings" related to attention [53]. In particular, brain–computer interfacing (BCI), specifically electroencephalography (EEG), efficiently captures immediate "first impression" responses at scale. Moreover, BCIs avoid the manual labor associated with "explicit methods" of feedback [10], which involve direct communication of intention through manual interaction. EEG constitutes implicit feedback, which relies only on the user's attention to stimuli and offers a streamlined and user-friendly approach to guiding generative models.

EEG-mediated information generation offers a novel human-in-the-loop approach using natural responses to visual stimuli. In this process, users' collective and implicit reactions

Manuscript received 1 February 2024; revised 3 May 2024; accepted 17 May 2024. Date of publication 18 June 2024; date of current version 8 October 2024. This work was supported in part by the Academy of Finland under Grant 322653, Grant 328875, Grant 336085, Grant 350323, and Grant 352915; in part by the Horizon 2020 FET Program of the European Union under Grant CHIST-ERA-20-BCI-001; in part by the Alfred Kordelin Foundation under Grant 230099; and in part by the Finnish Foundation for Technology Promotion under Grant 10168. This article was recommended by Associate Editor B. Lei. *(Corresponding authors: Carlos de la Torre-Ortiz; Tuukka Ruotsalo.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of Helsinki Ethical Review Board in the Humanities and Social and Behavioural Sciences.

Carlos de la Torre-Ortiz is with the Department of Computer Science, University of Helsinki, 00014 Helsinki, Finland (e-mail: ctorre@mailbox.org).

Michiel M. Spapé is with the Department of Psychology, University of Macau, Macau, China.

Niklas Ravaja is with the Department of Psychology and Logopedics, University of Helsinki, 00014 Helsinki, Finland.

Tuukka Ruotsalo is with the Department of Computer Science, University of Copenhagen, 1172 Copenhagen, Denmark, and also with the Department of Software Engineering, LUT University, 53850 Lappeenranta, Finland (e-mail: tuukka.ruotsalo@lut.fi).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2024.3406159.

Digital Object Identifier 10.1109/TCYB.2024.3406159

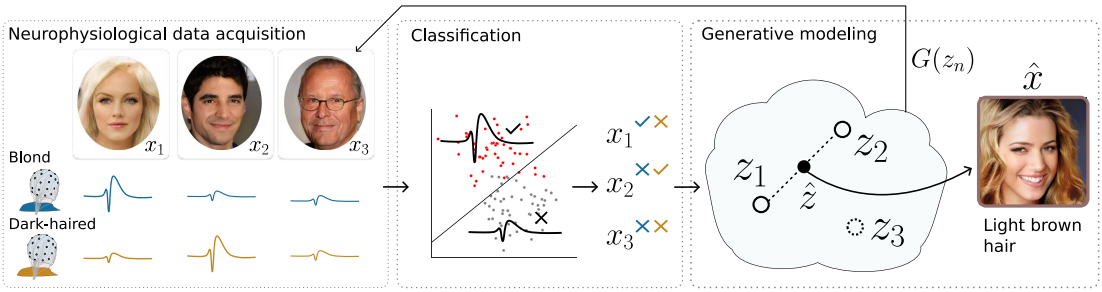

Fig. 1. EEG feedback from several individuals is captured as implicit feedback for salient VFs that individuals attend to (e.g., blond-haired and dark-haired images of faces). Then the recorded signals are classified to detect target and nontarget salient VFs in the images. Finally, the feedback is used to steer a generative adversarial neural network (GAN) to produce novel images of faces that represent the target salient features of the group of users (e.g., generate a novel image with intermediate and light brown hair features).

influence the output of generative models. These models integrate diverse feedback, including conflicting inputs that represent divergent preferences (e.g., combining "female" and "male" facial features), enabling collaborative input in generating computer-aided information.

In this context, we ask the following research questions:

RQ1: Can we infer the presence or absence of salient visual facial features of interest from EEG responses?

RQ2: Can implicit cross-subject EEG feedback be used to steer a generative model to produce novel images of faces capturing the mental targets of individuals?

RQ3: How do the number of users, the complexity of the image feature space, and potentially conflicting target VFs affect the quality of generated images of faces?

To answer the research questions, we demonstrate a proof-of-concept implementation of the methodology for sourcing cross-subject brain feedback, followed by generating new images of human faces with different VFs using a generative adversarial network (GAN). Data were collected in an EEG study where subjects viewed artificially generated images of human faces. Given the pioneering, early state of the technology, we design our experiment in a controlled setting: we instruct subjects to concentrate on a specific salient image feature intended to be in the output image, such as specific hair color, (normative) gender, age, or affective features, such as smiling. Machine learning models are then used to directly identify the presence or absence of such features in an image from brain responses. The corresponding latent representation of the image serves as feedback to modify the output of a generative model. This model is updated to produce entirely new images of faces that reflect VFs inferred to meet the subject's goals (Fig. 1).

We report a series of simulations to demonstrate the efficiency of an EEG-based image generation system based on input collected from a group of subjects in three image generation scenarios: 1) for generating images of faces containing a particular VF; 2) for producing images that combine several VFs simultaneously; and 3) finally, to resolve feedback with conflicting VFs, defined as those that are mutually exclusive, e.g., old and young.

To determine the efficiency of our cross-subject BCI for image generation, we evaluated them against two baselines: random feedback and explicit manual feedback. External assessors evaluated the output images generated using the different feedback models in a double-blind experiment. Our results show that the computer-generated images represent the target visual facial features nearly perfectly, are significantly higher quality than outputs resulting from a random process, and are comparable to those generated with explicit manual feedback.

The contributions of the research can be summarized as follows.

1) We present the first-of-its-kind methodology using implicit judgments inferred directly from cross-subject EEG responses to generate new visual information with artificial neural networks. Our approach contrasts with other generative BCI systems, which have been designed to only satisfy generation goals for a single subject.
2) We conducted experiments to integrate the assessments of visual information from a cohort of users via EEG, and generated novel images that satisfy a range of realistic yet controlled image generation tasks.
3) We investigated the effects of task complexity and the number of subjects providing EEG feedback on the quality of the generated images of human faces.
4) We evaluated the system's ability to reconcile simultaneous EEG feedback for VFs in conflicting semantic categories.

II. Background

Generative Modeling: Advances in generative models have made it possible to generate realistic and novel images. In particular, progressive training of generative adversarial neural networks [3], [15] has shown remarkable performance in image generation in several domains [30], [36]. While generative models have produced impressive output, their representations do not always correctly capture semantic features or concepts familiar to humans. In other words, these models learn latent representations, which may not directly match the human understanding of VFs. State-of-the-art GAN architectures, such as StyleGAN [29], address this challenge of matching more generalized features (the "style") of other images. While this approach aims to close the gap between human and model representation of VFs, it still does not control the generative process. The question of how human input can be considered in the parametrization of latent

representations to provide control over the generative process remains unsolved.

Recent research has also focused on gaining interactive control over generative models in creative tasks. For example, a generative model was used to complete a freely formed sketch of the intended final image. The generative model completed the picture with shapes and other features from the user's original sketch [38]. Research has also demonstrated approaches using text input for image synthesis. In this line of research, the system takes a text description as an input, and the generative model produces an image that fits the features of the text input [73]. Similarly, researchers have been using feedback on intermediate images to generate an image that better matches the user's intention in the subsequent iterations [68]. These approaches showcase the potential of generative models in visual information generation and other creative efforts. Nevertheless, they still rely on explicitly tailoring a specific model for a particular task. As the system undergoes supervised training, it learns to control the latent space by mapping interpolations between labeled features to the representation space. In addition, the visual information generation and manipulation task usually encompass an iterative process in which successive evaluations in the feedback loop bring user intentions and the visual perception of the output closer to each other [23]. Users can often provide diverse perception critiques of visual works, with advantages over peer feedback, such as avoiding an overly positive bias [65].

Neurophysiological Feedback: Online collaboration for design feedback has already benefited from including novel ways of interaction, such as video feedback [40]. Indeed, a more recent line of research has used BCI for affective computing [51], [52] or recognition tasks [10]: human cognition can be exploited in implicit tasks to obtain simple opinions or recognition signals. Thus, there is potential to utilize natural responses from users, where a system captures their immediate first impressions via neurophysiological data as passive input. This approach contrasts with other approaches in BCI, which aim to exert direct control of the computer by replacing motor movement (e.g., moving a mouse cursor) with a more limited range of applications [9].

Human cognition processes complex visual stimuli swiftly [62]. To an extent, categories and objects are also quickly recognized in EEG studies, which have facilitated emotion recognition [74] and motor imagery detection [7], [45]. However, object classification from EEG recordings ("EEG decoding") has proven exceedingly challenging [2], where only relationships between stimuli classes and neural responses have been successfully modeled, rather than decoding the category per se [25]. EEG signals can be used effectively to infer task relevance and reliably classify a limited set of binary and multiclass labels, enabling tasks, such as decoding user intention [72]. However, the performance is typically highest and most robust in a binary classification setting [8], distinguishing between target and nontarget outcomes. Indeed, direct interfacing with human cognition holds promise beyond simple recognition or image classification [27], [28], [55] tasks. Since such first impressions have a lasting effect on cognition, affect, and decision-making [64], detecting immediate stimulus evaluations directly from the brain can provide optimized estimates for many tasks that benefit from *rapid and binary* human feedback on visual content.

Connecting EEG With GANs: Electrophysiological data can provide these immediate stimulus evaluations by detecting the degree to which displayed features implicitly match the target features that a user has in mind. For example, the event-related potential (ERP) is an electroencephalographic metric that quantifies brain activity synchronized to external events, such as the onset of images. Therefore, EEG can provide feedback on perceived visual information.

While current EEG-based approaches may not reveal complex goals or high-level outcomes, it is possible to detect whether an image contains VFs relevant to the user. In particular, the P3 is a late parietal positivity in the EEG, which is well-established in cognitive neuroscience to be particularly evoked by infrequent, task-relevant stimuli [11]. The P3 may therefore provide an implicit biomarker for enhanced processing of target stimuli, such as VFs, even in the absence of overt, physical responses. Another line of research has approached feedback on brain signals using the entire ERP, which includes the P3 [5], [12], possibly responding to more complex activity patterns. Recent work has aimed at connecting GANs with neurophysiological signals [10], [18], [58], [67]. For other research in this direction see [31], [44], [57], [63]; however, later replication work has shown that they exploit the block structure of their experimental design, and therefore the results have been questioned [1], [4], [33]. Indeed, multiple features of the EEG signal are naturally affected by the temporal order of the block design; therefore, adding all target stimuli at the end of the experimental blocks produces an artificial positive classification. In reality, this is a confound with the natural temporal properties of EEG rather than stimulus-related activity [2], [33]. In contrast, we carefully control for these effects in our work with a randomized design that follows the "oddball" EEG paradigm [60]. Our experimental design uses complete randomization of both target and nontarget stimuli classes within the same experimental block.

As a result, here we put forward a first-of-its-kind methodology that enables a direct interface between implicit EEG feedback from many individuals and a generative model of images. Reactions from a cohort of users are obtained directly from their brain responses, which are then used as collaborative brain feedback to generate new images of faces that match the target VFs by the individuals.

III. Methodology

This study employs a methodology consisting of five phases.

1) We collected neurophysiological EEG data from 30 subjects in response to an image presentation task while keeping a mental target (facial feature) in mind.
2) We trained a classifier that maps each image onto the target and nontarget classes using the brain signals.

3) We used the classifier feedback to update latent estimates with positively classified stimuli within the VF space learned by a GAN.
4) We simulate a group image generation task, which used updates from several individuals to update the GAN estimates and then generate visual information from the GAN along image generation goals defined as having single, multiple, or conflicting VFs.
5) Finally, we evaluated the performance of the EEG-based group image generation process against control processes based on random and explicit feedback.

A. Neurophysiological Data Acquisition

Thirty study subjects (13 female, 17 male; self-reported) volunteered to participate in the neurophysiological experiment. They were, on average, 28.2 (SD = 7.1) years old, with no restrictions on age range eligibility, although participation was advertised to a university population. They all had normal or corrected-to-normal vision and no history of using neuropharmaceuticals. The ethical review board in humanities and social and behavioral sciences of the University of Helsinki approved the study. It complies with the protocols laid out by the declaration of Helsinki. Subjects were fully informed of the study and their rights, including the right to withdraw at any point, and signed a consent form before participating. They received cinema vouchers for their time and efforts.

Stimuli and Apparatus: We used a pretrained GAN architecture[1] based on celebrity photographs to generate stimuli [30]. It produced an initial set of ca. 10 000 stimuli by random sampling latent vectors with a multivariate Gaussian distribution across the 512-D space. These were then manually inspected to filter out visual artifacts, following which they were categorized into eight discrete VFs: blond- and dark-haired; (normative) female and male; nonsmiling and smiling; and old and young. These features were selected as we expected all study subjects to understand and recognize them easily, and they were well-represented in the model. Of these, 260 stimuli on average per feature were selected for use in the study, such that each subject saw more than 2000 different images of faces in the experiment. To normalize the stimuli across unrelated dimensions, we applied an ellipse cut-out frame such that mainly the foreground containing the central face was showing (Fig. 2). Furthermore, to improve timing accuracy during image presentation, we down-sampled the 1024 × 1024 images to half this resolution.

We presented the images using a rapid serial visual presentation paradigm (RSVP) developed in E-Prime 3 [59] to optimize the presentation timing and physiological recording synchronization. The setup used a 24" LCD screen with a resolution of 1920×1080 @ 60 Hz situated around 60 cm from the subject to display stimuli. A BrainProducts QuickAmp USB 32 amplifier digitized the data at 1,000 Hz from 32 passive silver/silver-chloride electrodes placed on an EasyCap system to ensure optimal equidistant placement at sites FP1, FP2, F7, F3, Fz, F4, F8, FT9, FC5, FC1, FC2, FC6, FT10, T7, C3, Cz, C4, T8, TP9, CP5, CP1, CP2, CP6, TP10, P7, P3, Pz, P4, P8, O1, O2, and Iz, with AFz as ground.

[1]https://github.com/tkarras/progressive_growing_of_gans

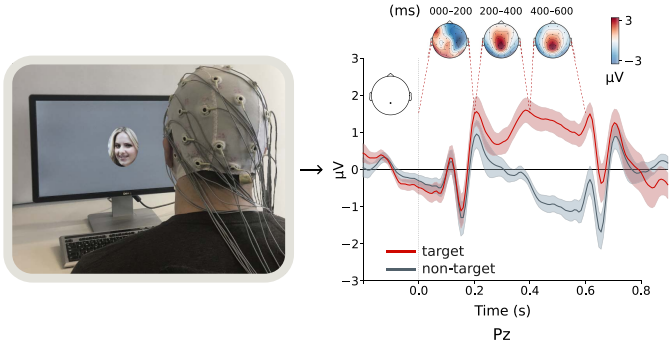

Fig. 2. (Left) Setup of the neurophysiological data acquisition in which EEG is recorded in response to viewing images of faces. (Right) Average ERP plots for all subjects are shown with the standard deviation for the Pz channel in the −200–800-ms window and topographic maps of the averaged scalp electric potential within 0–200, 200–400, and 400–600-ms post-stimulus windows.

Procedure: Following hardware setup and signing of the informed consent, study subjects started the data acquisition session by observing images of faces in RSVP sequences. They were tasked to keep a mental target based on one of our eight selected VFs (e.g., blond-haired, dark-haired, female, male) in mind throughout each series of 70 images. Subjects simply concentrated on the images that presented the target feature while limiting unnecessary mental or physical activity. For example, subjects concentrated on images of blond individuals in the task involving "blond-haired" features as a target. Images were divided into those labeled as target (∼30%, e.g., blond-haired) and nontarget (∼70%, e.g., dark-haired) categories, where target images were intentionally under-represented and spaced to evoke P3 potentials. We expect this ERP to be the main contributor to signal classification performance for downstream analyses, and these potentials are known to be amplified by relatively rare and task-relevant stimuli.

Each trial included 70 images, with 20 labeled as target and 50 as nontarget stimuli. We presented images at 500-ms intervals without an interstimulus interval, and each image was displayed once per trial. Following each trial, a 500-ms gray mask appeared, leading to a self-terminated break. For instance, in a "blond-haired" block, each subject viewed 80 unique images of blond individuals and 200 dark-haired ones, with the reverse setup for a "dark-haired" task. The order of tasks was counter-balanced and randomized among subjects. A complete session involved displaying 2240 images, spanning approximately one hour, excluding time for setup and instructions.

EEG Preprocessing: EEG data was preprocessed to reduce noise and improve signal quality. It included band-pass filtering, time-locking to stimulus onset, subtraction of the average prestimulus value to correct for baseline, and removal of artifacts via a thresholding technique. First, continuous EEG data were band-pass filtered between 0.2 and 35.0 Hz to reduce slow drifts and high-frequency noise. Then, they were time-locked to stimulus onsets and segmented into 1100-ms EEG epochs, including 200 ms before the stimulus [19], [26].

Finally, we removed EEG epochs containing strong artifacts related to eye movements, blinks, and muscle activity from the dataset. In particular, we used a thresholding technique that relied on individualized maximum absolute voltage cut-offs: we first selected the 0–50-ms post-stimulus window of the first 2000 epochs. Then, we picked the 200th highest-absolute value per subject and constrained (clipped) these thresholds to values between 10 and 80 μV. As a result, this approach tagged 11.36% of epochs as artifactual on average and removed them from the analysis.

Following preprocessing, we recorded each subject's data in a $e \times c \times t$ tensor (e epochs, c channels, and sampled points in time t). Then, we divided each epoch into 7 equidistant time windows per channel on the 50–800-ms post-stimulus period, based on previous studies [26] and a liberal estimate of the known time window of the P3 [19]. Finally, we averaged all data within each time window for each time frame, with all existing channels and time frames combined to produce spatiotemporal vectors.

B. Classification

Classification Setup: Randomly sampled images were shown to the study subjects during the neurophysiology data acquisition while their EEG signal was registered, simulating an independent calibration step across subjects. In sum, the classification step aimed to assign a target/nontarget class label to an ERP independent from the analyzed VF category. We trained regularized linear discriminant analysis (LDA) single-trial ERP classifiers for each subject and VF. In particular, we used the Scikit-learn implementation with the least squares solution combined with automatic shrinkage using the Ledoit-Wolf lemma [5], [47]. To minimize false positives, we considered only predictions with a confidence score higher than 0.7 for the target class based on the training performance.

Data Preprocessing and Classifier Input: Models processed input vectors composed of time series data concatenated over multiple epochs, where each series represented the voltages recorded across all channels. We assigned the manual binary class membership labels (target or nontarget) to the vectorized representations of the ERPs, which together were used as input to train the classifiers. Thus, the evoked potentials should be positively classified when looking at an image containing the target VF (e.g., "Look for blond hair" → image of a blond-haired individual). Otherwise (e.g., "Look for blond hair" → image of a dark-haired individual), the stimulus and associated ERP should be classified as nontarget.

Classification Tasks and Dataset Partitioning: In total, we trained eight independent classifiers, one per visual task, ensuring that they learned from all tasks and, at the same time, were agnostic to specific VFs. Each classifier used data collected during all tasks, excluding the target class and its opposite (e.g., blond- and dark-haired) as the training set and the target class as the test set (e.g., blond-haired) as depicted in Fig. 4. . For example, a classifier for the "young" VF was trained using data from the "blond-haired," "dark-haired," "female," "male," "nonsmiling," and "smiling" features, but not from the "young" and "old" tasks. Therefore, our approach aimed for a rigorous data split by also excluding from training any VFs that belong to the same broader category as the target class, e.g., age. As a result, we ensure that no task-related features unrelated to the stimuli class would affect the classifier performance [33]. Such exclusion ensures that the classification performance was not inadvertently enhanced by learning to classify the opposite VF in the same stimuli class. As a result, we split the data into a training set (75% of the samples, average 1497.52 per visual task and subject) and a test set (12.5% of the samples, average 249.59) based on the VFs and subject. The remaining 12.5% of the samples corresponding to the opposite VF were not used.

C. Generative Modeling

We first recorded the neural signals in response to the subject's perception of a generated image, categorizing them as "target" or "nontarget." The system then used the task membership label, together with the corresponding latent representation of the image, and combined the feedback of many individuals to generate an output image. Thus, the GAN uses a neuroadaptive approach similar to [26], [71], such that feedback updates the estimate of a target image in the generative model's latent space. Both the classification and generative steps utilize the same eight single VFs.

Generative Model: The generative model (GAN [15]) provides a mapping between the latent space and the stimulus (image) space. GANs are a type of artificial neural network composed of a discriminator D and a generator G during training. G attempts to achieve generative performance to output images that cannot be told apart from training examples. By comparison, D requires distinguishing whether input images belong to the training dataset or are a G falsification. Consequently, G and D are involved in adversarial training until D cannot reliably discriminate between training samples and G output. We can exclude D at this stage and assume G is fully trained. As a result, the generative model has a latent representation of training data.

The generative model we used was pretrained with images of celebrity faces (CelebA-HQ data set) [30]. This dataset is a higher-resolution variant of the CelebA dataset [37]. The output of the GAN are 1024 × 1024 px realistic images of faces generated from 512-D latent vectors. Therefore, the model provides a mapping of $G : Z \to X$, such as $G(z_n) = x_n$, where $z_n \in Z$ is a point in a latent space Z and $x_n \in X$ is graphical information perceivable by humans (images of faces). To that end, the generator creates images $(x_0, \ldots, x_n)$ from vectors $(z_0, \ldots, z_n)$ taken from the latent space.

Feedback and Model Updating: The goal of our framework is to steer the generator G within the latent space Z to find a point $\hat{z}$ where the intended VFs are represented. This point is such that $G(\hat{z}) = \hat{x}$ aligns closely with the group's collective mental VF(s). As more subjects contribute their feedback, latent vectors classified as "target" $\hat{z}_n$ progressively refine the estimation of $\hat{z}$: each successive iteration $G(\hat{z}_n) = \hat{x}_n$ is expected to reflect the intended VFs more accurately than the previous $G(\hat{z}_{n-1})$.

In detail, the technique for updating the generative model is as follows. We sampled a stimulus image from a latent vector from the GAN with an increasing the number of subjects. The sampling strategy ensured that nontarget vectors, those corresponding to the opposing VF, are over-represented compared to the target class. This requirement enabled images with desired features to evoke intensified brain potentials when presented with those without them. Moreover, this is a realistic scenario in which target VFs are not favoring the output by a majority vote effect on the classification step.

Responses update the latent estimation with a variant of the well-known Rocchio algorithm [50], chosen as the simplest vector interpolation. That is, the latent vectors of images corresponding to target classifications update the GAN estimate, so the generative output captures the group's opinion. We determine the average of the image vectors as those positively classified and designate z_{avg} for the corresponding vector. Then we update the z_n vector by having $z_{n+1} = z_n + z_{\text{avg}}$, where a new randomly chosen subject provides z_{n+1}, but without any sequential dependencies between each z_n. The z_n point can now shift step-wise in Z and finally reach the $\hat{z}$ position as positive classifications from new users updating the feedback pool. This final $\hat{z}$ produces the $G(\hat{z})$ image with the target VFs in consensus with the user group.

D. Composing User Groups With Divergent Goals

Combined Visual Feature Simulation: We simulated the combination of VFs by integrating input feedback corresponding to the individual VFs. Consequently, we collected the appropriate cross-subject feedback sequentially from multiple VFs in the form of latent vectors z. For example, for a combined image with "male" and "old" VFs, positively classified vectors from the "male" class were used to update the latent space, followed by vectors from the "old" class. In these simulations, we omitted combinations that would include opposing VFs (e.g., "smiling" + "nonsmiling"), as these features conflict with each other.

For example, if we simulated study subjects contributing feedback for a combination of VFs, e.g., half of them would focus on "smiling" faces and the other half on "blond-haired" faces. In this case, we would expect the generated image to be the face of a smiling blond person.

Conflicting Visual Feature Simulation: In an experiment independent of the previous, we generated image pairs for each set of two opposite, or conflicting, VFs with EEG feedback from all subjects. Such pairs of mutually incompatible features were formed from our set of visual tasks, for example, simultaneous feedback from both "smiling" and "nonsmiling" VFs. Therefore, we simulated conflicting feedback by updating the generative model with equal brain feedback from each opposing VF. In the previous example, we would select 15 randomly sampled and positively classified latent vectors for the "smiling" VF and 15 randomly sampled and positively classified latent vectors for the "nonsmiling" VF. The resulting mean vector was used as input for the generative model, resulting in one output image. The process yields sets of three images: two corresponding to the opposing VFs and one representing the conflicting scenario. Therefore, our total of eight VFs result in four sets of images, as seen in Fig. 6.

E. Evaluation

We employ different evaluation methods to measure the classification performance and image generation quality, which we explain below.

Evaluation of Classification Performance: Classifier performance was measured by area under the ROC curve (AUC) and validated by a permutation test. In the permutation test, we obtained permutation-based p-values by contrasting task classifiers' AUC scores with those of classifiers trained with randomly permutated class labels [43], accounting for unbalanced classes. We determined a minimum theoretical p-value of 0.01, performing $k = 100$ permutations per subject [44]. Finally, we computed the AUC of predicted labels for the target VF (test data) against ground truth labels manually assigned for each stimulus image (manual VF annotations).

Control Models and Evaluation Protocol: We introduced two additional feedback baselines to assess the semantically salient VFs relevant to the task in generated images. First, a random process was designed to represent the empirical lower bound of our approach's performance. In this model, the classifier's labels assigned to each stimulus image were randomly shuffled, distorting the EEG-based target versus nontarget label and image pairs. Second, an explicit feedback model was used to represent an empirical upper-performance limit. Here, each image was assigned its manual ground truth label for the same target versus nontarget classification, effectively bypassing brain-based inputs. Therefore, positive feedback was assigned to images that correctly match the VFs set as the target for each task.

Generating Images That Combine Visual Features: We generated and evaluated images combining 1, 2, 3, and 4 VFs of faces, covering all 4×4 feature combinations. Each subject contributed one VF to the global target image generation goal (e.g., Subject A contributed "blond-haired" and Subject B contributed "male" for "blond male"). The model then combined feedback from single VF targets into a cross-subject output. Fig. 5 shows images generated by combining several VFs using the brain, random and manual feedback models, and increasing the number of subjects. Individual contributions were used as feedback to update vectors in the GAN latent space, outputting an image that combined all target VFs.

Two external annotators evaluated the generated images—one self-reported as female and 57 years old, and the other as male and 25. They had not participated in the neurophysiology experiment. The images were displayed together with their corresponding task (VF description), and one evaluator rated them on a discrete scale: "no match" (0), "partial match" (0.5), and "total match" (1). In the evaluation, the order of the images was randomized, meaning that the annotator did not know which of the three processes produced each image. Therefore, the process was blind, and it objectively determined the performance of the processes. To ensure the reliability of the annotations, another evaluator assessed 100 randomly

selected images. The resulting Cohen's Kappa shows a high-inter-rater agreement between the annotators ($\kappa = 0.76$).

Evaluation of Generated Images With Conflicting Goals: Additional external assessors evaluated the generative outcome of the conflicting visual feedback simulation in another annotation study that included ten independent annotators (mean = 41.6, SD = 13.6 years old; two female, eight male, self-reported). A larger pool of evaluators was used because the images were intentionally generated with conflicting feedback and their match with the categories was more nuanced. This study design ensured that annotations would represent a more general opinion and result in a more reliable estimate of the quality of the generated images. In the evaluation, we presented the annotators with three images: two with the opposite tasks and one combining these conflicting tasks. For example, one image with feedback on "young," one image with feedback on "old," and one image with conflicting feedback on both "young" and "old" (same amount for each feature).

The annotators then evaluated the image with conflicting feedback on a continuous scale from 0 to 1. For example, in the "young," "young-old," and "old" image set, a score of 0 would mean conflicting output completely matching the "young" image features. A score of 0.5 would mean a perfect intermediate between the "old" and "young" VFs. A score of 1 would mean conflicting output completely matching the shown "old" image features. Each evaluator assessed all four conflicting images.

IV. Results

A. Neurophysiological Findings

A dissociation between target and nontarget stimuli may be observed from ca. 250 ms after onset, as shown in Fig. 2. We observed this over centro-parietal sites and gradually grew in magnitude (maximum difference of 2.36 μV at 464 ms, $t(30) = 11.80, p < 0.00001$) until well after the onset of the subsequent stimulus (at 500 ms). Brain responses to target images of faces showed a third positive peak with a latency of ca. 380-ms absent in brain responses to nontarget images. We observed a similar positivity across tasks within the 250–450-ms range. Thus, the latency, topography, and task dependence suggested that the effect of target stimulus detection reflected a P3 response [48], which is expected to be seen in such an oddball detection task.

B. Classification Performance

The training procedure was in the milliseconds range for all subjects and visual tasks ($N = 240$) with mean 61.15 ms (SD = 82.31, min = 45.82, max = 1329.46, and $N = 240$). To evaluate the classifiers, we computed the mean AUC, precision, recall, and F1 score over the study subjects. Table I shows the average performance per VF. Throughout, the median classifier AUC was above 0.7, indicating acceptable classification performance [41]. The high precision and moderate recall score in Table I indicate that models are conservative to avoid false positives, as expected from the selected confidence score.

TABLE I
ROC AUC, Precision, Recall, and F1 Scores for Individual and Aggregated (Average) Visual Stimuli Tasks (Blond-Haired, Dark-Haired, Female, Male, NonSmiling, Smiling, Old, and Young)

	BH	DH	F	M	NS	S	O	Y	*A*
AUC	0.83	0.80	0.85	0.80	0.78	0.81	0.73	0.74	0.79
Prec.	0.79	0.76	0.79	0.75	0.70	0.76	0.65	0.65	0.73
Rec.	0.31	0.28	0.37	0.29	0.28	0.32	0.28	0.29	0.30
F1	0.43	0.40	0.49	0.41	0.39	0.44	0.38	0.39	0.42

C. Quality of Generated Images

The image quality convergence was evaluated by increasing the number of subjects, as shown in Fig. 3. In general, adding feedback from an increasing number of subjects enhanced the performance of the generative model across all VF combinations. Performance increased steadily with EEG and explicit feedback as a function of the number of subjects, showing that increased subject count provided additional feedback for the generative model, improving the latent space estimate for the target VFs. Increasing the task complexity (the number of concurrent VFs) entailed lower quality and convergence speed. Nonetheless, concerning the user evaluation of image generation quality, using several target features simultaneously also increased the likelihood of producing a partial match. Contrarily, due to a class imbalance favoring the negative class, the performance of the random baseline based on label permutation decreased as the number of subjects increased.

Table II shows data for all combinations of VFs, significance, and improvement in output quality across feedback models. We computed the statistical significance of quality scores via a t-test, and Bonferroni corrected for multiple testing. When all feedback was available to the generative model, we observed image generation performance near parity between the brain and explicit feedback models. The quality of the generated images was comparable to those resulting from explicit manual feedback.

D. Quality of Generated Images With Conflicting Goals

We evaluated the generative outcome in case of conflicting image generation goals. The system generated images of faces with equal brain feedback for opposing VFs (e.g., "smiling" and "nonsmiling") and resolved cross-subject disagreement by displaying intermediate VFs (Fig. 6). Conflicting feedback for "blond-female" and "smiling-nonsmiling" output tended to produce an output image close to a perfect intermediate between the two extreme features. On the other hand, the "young-old" simulation tended to produce images with features closer to "young." The output of a conflicting "female-male" tended to produce a female, but with facial features tending androgynous features, as indicated by the disagreement from evaluators. With previous observations of generative output per study subject, some VFs tended to be over-represented, e.g., faces with "female" or "young" VFs were generated more often than "male" or "old" features, probably as those features are correlated in the GAN's training images and thus in its learned space.

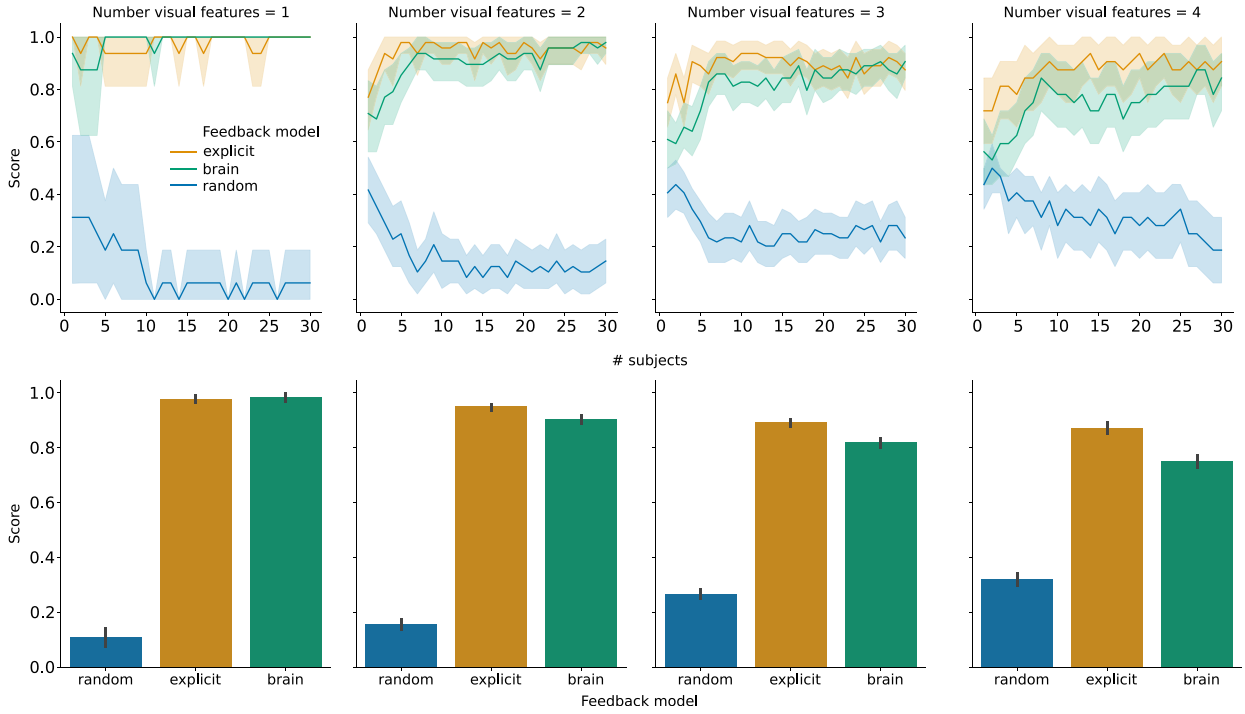

Fig. 3. Image quality across VF combinations (1 to 4 as columns) and feedback models: brain (green), random (blue), and explicit (orange). (Top) Mean quality score and its 95% confidence interval as the number of subjects increases. (Bottom) Mean quality score achieved by the maximum number of subjects in the experiment ($N = 30$).

TABLE II
QUALITY OF THE GENERATIVE OUTPUT FOR DIFFERENT AMOUNTS OF VF COMBINATIONS (#VF) AND FEEDBACK MODELS ($N = 30$; R : RANDOM, B : BRAIN, E : EXPLICIT; FB : FEEDBACK). MEAN VALUES RANGE FROM 0 TO 1 (HIGHER IS BETTER) AND ARE SHOWN WITH STANDARD DEVIATION, THE PERCENTAGE DIFFERENCE IN THE GENERATIVE OUTPUT BETWEEN THE BASELINE AND FEEDBACK METHODS (Δ), BONFERRONI CORRECTED p-VALUES, AND RESPECTIVE T-STATISTICS BETWEEN THE CONDITIONS

# VF	R FB $\pm\sigma$	B FB $\pm\sigma$	E FB $\pm\sigma$	(T) B/R	(T) B/R	Δ B/R	Δ B/E
1	0.19 ± 0.18	0.84 ± 0.00	0.91 ± 0.00	21.03	-0.39	$+93.75^{***}$	$+0.00$
2	0.06 ± 0.23	1.00 ± 0.10	1.00 ± 0.14	17.64	-1.29	$+83.33^{***}$	$+2.08$
3	0.15 ± 0.25	0.98 ± 0.20	0.96 ± 0.22	13.00	-1.26	$+67.19^{***}$	$+3.12$
4	0.23 ± 0.25	0.91 ± 0.24	0.87 ± 0.20	8.57	-2.39	$+65.62^{***}$	-6.25

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

V. DISCUSSION AND CONCLUSION

We presented an implicit BCI system for image generation. To this end, we reported a proof-of-concept system and experiments in which an individual or a group of subjects can provide implicit feedback to a generative model producing novel images of faces matching cross-subject mental targets. The approach demonstrates computer-generated visual information from EEG feedback without explicit human input; human subjects only pay attention to VFs relevant to their target image generation goal.

A. Answers to Research Questions

We asked three research questions to study whether BCI for visual information generation is possible and how it performs compared to random and explicit feedback. Here, we discuss the results accordingly.

RQ1: Can we infer the presence or absence of salient visual facial features of interest from EEG responses?

Our findings demonstrate that subjects recognize specific VFs and that the brain signals evoked by this recognition are consistent with previously reported neurophysiological findings [12]. The machine learning experiments show that the feedback can be reliably decoded and that the performance of the single-trial ERP classifiers trained using brain signals significantly outperforms random classifiers and shows an average classification performance (AUC) of above 0.7. Although the models are personalized and trained individually for each study subject, thus requiring personalized calibration, the average performance of the group of subjects is remarkably high.

RQ2: Can implicit cross-subject EEG feedback be used to steer a generative model to produce novel images of faces capturing the mental targets of individuals?

The approach was shown to control the generative process toward target VFs adequately. The quality of images of faces generated with implicit EEG-based feedback is comparable to manual selection and significantly better than a random

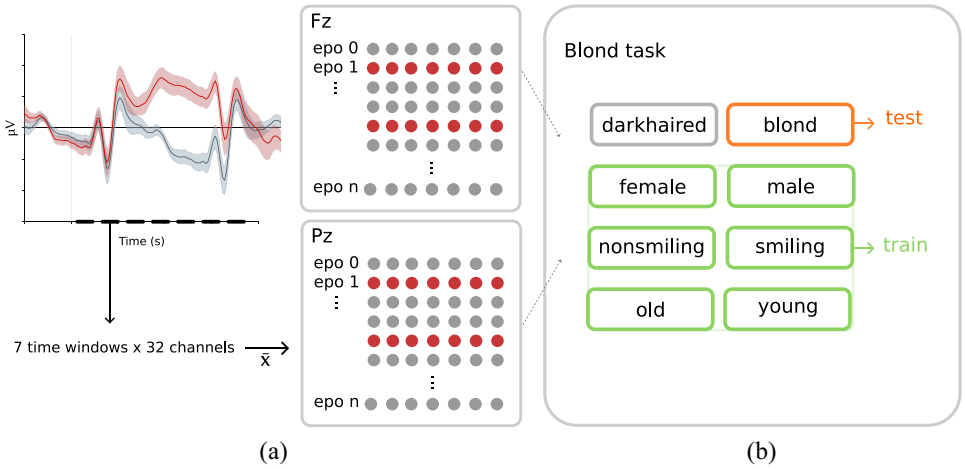

Fig. 4. (a) EEG epoch feature engineering into spatiotemporal vectors in response to a target (red) or nontarget (gray) stimulus. Feature vectors are separated according to the VF task they were recorded in. (b) Training and test data splitting for the example "blond-haired" task. The training set includes all data but the target visual task and its opposite. The test set includes only data for the target visual task. This conservative setup avoids confounds of training on brain responses of the same visual category.

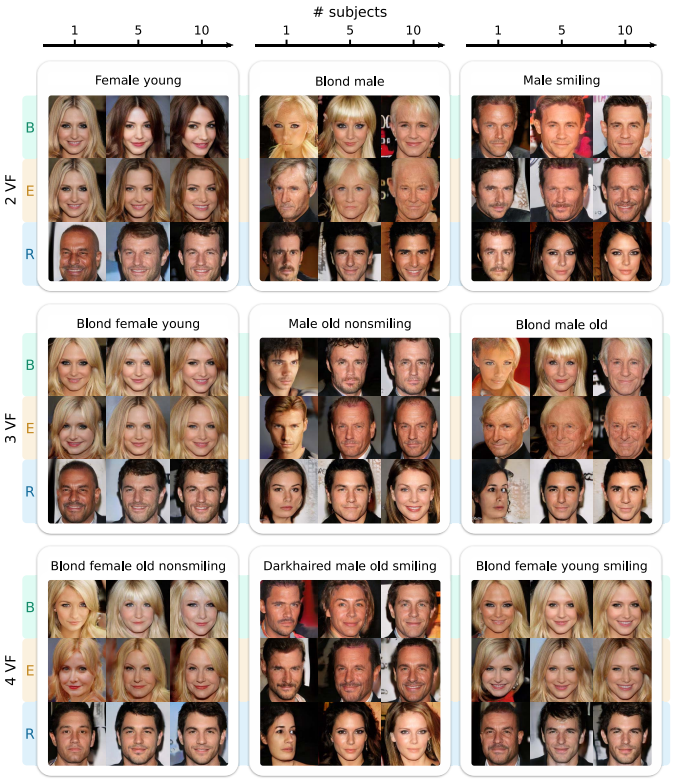

Fig. 5. Examples of generated images of faces combining 2, 3, and 4 VFs (VF) using the three feedback models: brain (B), explicit (E), and random (R). Generative performance increases with the number of subjects, especially for more challenging tasks, such as combining more VF or VF over-represented in the model's training data. Brain and explicit feedback produce images that match the group's goals.

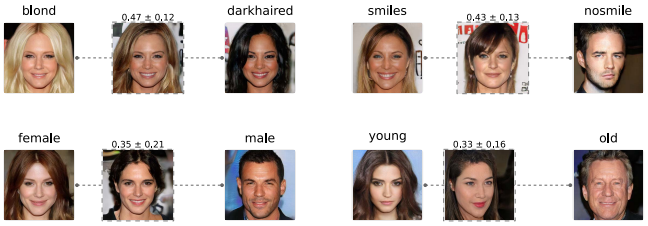

Fig. 6. Image generation output with conflicting VFs. Blocks display a pair of opposing VFs generated from all subjects. The middle image was generated with sampled half feedback pool from the opposing visual tasks. Results are shown as the mean and standard deviation of similarity to the edge examples (for a VF—middle image; 0: equivalent to the respective left image; and 1: equivalent to the respective right image).

baseline, including scenarios in which multiple VFs, are combined.

RQ3: How do the number of users, the complexity of the image feature space, and potentially conflicting target VFs affect the quality of generated images of faces?

The number of users is shown to affect the quality of the generated images significantly. For models trained with multiple users, as few as ten subjects are enough to provide high-quality generative output that is not significantly different from the output obtained with 30 subjects. This suggests we can effectively capture diverse mental targets in creative tasks through implicit feedback, even with as few as ten subjects contributing to the process.

Our results also show that the results generalize to image generation scenarios with complex multidimensional goals. While study subjects focus on individual VFs (e.g., "smile"), the results from simulations in which different VFs are combined (e.g., creating a smiling face of a dark-haired young male) show that the methodology generalizes to multifeature scenarios. As expected, the VF space's complexity affects the generated images' quality, but the general image quality remains very high.

The results indicate that our approach can generate images of faces despite conflicting goals and tolerates conflicting cross-subject feedback. However, balancing the contributions of individual subjects in these scenarios remains challenging. Some conflicting goals, such as blond- and dark-haired, were

better accounted for than others, such as young and old. Our Rocchio feedback approach may have been too simple to capture the scale invariance in latent spaces, resulting in imbalanced feedback and imbalanced presence of VFs in the output images. Another possible explanation is that the latent GAN space does not capture the dimensions equally and leads to the collapse on some of the target features. Despite these shortcomings in combining conflicting feedback, our results show that the models adequately reflect the user's mental targets and can lead to high-quality output even in more complex image generation scenarios.

B. Limitations and Future Work

Stimulus Selection and the P3: Our experiments focus specifically on the VFs of images, particularly of human faces, which leverages our natural human ability to discern and categorize facial features. Furthermore, it allows the investigation of complex visual categories like "old" or "male" in a context that study subjects can intuitively understand. In addition, we base our approach on the subcomponents of ERPs that reflect the target relevance of features appearing in stimuli. Therefore, our approach is not dependent on any particular stimulus feature, and the presented methodology has the potential to work on different datasets, image types, and tasks. Indeed, the targeted ERP components are known to occur across visual, auditory, tactile, and even olfactory modalities [42], [61]. However, we did not experiment with other forms of image generation beyond facial features. Therefore, we cannot entirely exclude the possibility that other datasets and generative models might lead to differences in performance, which is an exciting avenue for future work.

Also, for experimental reasons so that the outputs generated in response to feedback could be objectively evaluated, our methodology used VFs that are a priori objective, easy to recognize by users, and with an assigned opposite feature (e.g., "smiling" versus "nonsmiling" rather than "liking" versus "disliking"). However, there is still a subjective component in interpreting some of these. For example, our perception of "old" versus "young" can change as we age. Even with manual explicit feedback, it can prevent reaching maximum evaluated performance ratings in the image generation task.

It is noteworthy that the system does not decode the identity of the stimulus from the brain signal. Instead, it detects whether the VFs of interest are salient in a particular stimulus or not, and thus, we expect the P3 to be sensitive to the general saliency of features in our BCI design. Therefore, the methodology should generalize to other types of visual stimuli and even other modalities. As part of our preanalysis, our results report that feature relevance indeed shows an ERP oddball effect with a clearly amplified P3 component. However, the classifiers use the entire ERP and do not necessarily rely only on the P3, but we expect the P3 to contribute notably to the models.

Classification and Generative Modeling: Concerning stimuli selection, EEG responses to the negative class were over-represented due to the design of the neurophysiological data acquisition. A random classifier with such class imbalance will easily learn to predict the majority class. Consequently, the feedback would cause the latent vector to diverge from the target, and, as a result, the generative model would be more likely to generate images that match the opposite task, e.g., "dark-haired" features in the "blond" task, thus decreasing image quality with random feedback. As a result, correctly classifying the positive target class with such imbalance makes the task even more difficult, the expected random performance lower, and our final image generation task more difficult. Despite this imbalance, we observe high-classification performance and image generation performance comparable to explicit feedback.

Demographic factors, such as the user's age and gender, could influence the accuracy of decoding visual preferences and downstream generative modeling performance. However, we believe that these factors do not diminish the significance of our findings. In addition, while our approach shows promise for broader applications across various generative models, such as variational autoencoders [32] or diffusion models [22], this initial work did not extend to such comparative analyses. However, for our contribution, the particular choice of generative model is secondary, and we anticipate that it would lead to subtle differences in performance only. Exploring additional sources of variability and assessing the generalizability of our methodology across different generative modeling techniques constitute a key direction for future research.

Usability and Deployment in the Wild: Our approach leverages cross-subject implicit feedback to produce novel visual information, also improving the EEG signal-to-noise ratio. While our approach was implemented as a proof of concept, we relied on a separate experiment to collect data and analyze image generation results. Our methodology allows online adaptation but was not implemented for the present models and stimuli selection. In practice, the time for classifier training and updating the system based on feedback was negligible, whether the aggregated feedback contained multiple and/or conflicting VFs. The most significant software bottlenecks were related to EEG preprocessing and image generation, reflecting current technological constraints rather than our methodology. Therefore, an online scenario with an adaptive stimuli sampling strategy is subject to future work but has been demonstrated to be feasible in the simulations.

The comfort, calibration, and general acceptability of wearable sensors could limit an online experiment to work outside of laboratory environments. However, we expect the hardware to become more comfortable, wireless [6], [21], [69], and easier to calibrate and wear due to the adoption of dry electrodes. Moreover, we foresee further advantages from lines of research, such as self-calibrating BCI systems [16] and classifier-free generative modeling [66]. These advancements could make use of more nuanced EEG responses, smoothly altering the VFs of GAN's outputs.

Applications of EEG-Based Interaction Modalities: We already anticipate some immediate advantages of this interaction modality compared to traditional interfaces. Our

method is better suited for collecting fast and intuitive evaluative feedback, not scenarios with prolonged complex tasks or fine-grained control. We also expect that in our approach the mental effort alone will be less demanding than that of the motoric and mental load of manual methods. However, we did not record self-reported measures to compare mental and physical load, such as [20], [39]. Thus, we envision a future in which similar EEG-based methods will coexist with traditional interfaces, and the present study should be considered the first-of-its-kind proof of concept. Similarly, our system has unique advantages: it detects feature presence based on first impressions and allows fully implicit interaction. However, a detailed comparison to other input methods requires further experimentation and investigation of usability in more realistic scenarios.

Collectively generating images from brain signals enables various applications to create personalized visual experiences. In such scenarios, preferences from many users can be combined into a final output that captures both diversity and agreement. For example, users from a specific demographic could provide feedback and shape an image toward a brand feeling, as implicitly understood by such a crowd.

Ethical Considerations: Finally, we identified two main ethical challenges: 1) due to a bias in the generative model and 2) related to the use of BCIs. First, as shown in Figs. 5 and 6, the output of the generative model shows a predisposition to generate certain features since the model was pretrained on a database of images of celebrities. Likewise, specifying or controlling other attributes was out of the scope of the present study but should be considered in follow-up research. Also, we advertised and enrolled study subjects from a university population in Finland. A more diverse group of subjects should be considered for future studies.

Second, an important concern arises from the nature of BCI systems. Despite the current technological limitations for real-world applications, the regular or pervasive use of wearables could raise serious privacy concerns and potential misuse of the technology [54]. While these technologies are not inherently harmful, collecting physiological data introduces the potential for misuse, such as accumulating large datasets of sensitive signals, fingerprinting, and inappropriate consent procedures, among other issues [24]. Furthermore, the preambles of the European Parliament legislative resolution for the Artificial Intelligence Act [46] identify brain–machine interfaces as potential enablers of deceptive content that could challenge individuals' autonomy through a high degree of personalization of content. Given this significant potential for misuse, there is a need for research and further development of guidelines that ensure the open and democratized development of these technologies.

C. Conclusion

To the best of our knowledge, we present the first-of-its-kind cross-subject system for image generation using implicit feedback. Our system uniquely utilizes implicit feedback directly from the brain, effectively integrating diverse and potentially conflicting goals from multiple individuals. Our approach takes advantage of the very first, immediate reactions occurring within less than a second, from which users form their opinion, by directly capturing this effect from neurophysiological signals. Furthermore, we close the loop by producing novel visual information that meets people's intuitive, visceral responses simply by observing their brain potentials. We exemplify this proof of concept work with the case of automated image generation in several user studies and evaluations. Our approach successfully aggregates the natural reactions of a cohort of subjects as evaluative feedback and generates new visual information matching those, achieving remarkable performance across tasks. With this, we envision an intriguing future in which users can make images matching their mental targets aided by responses directly captured from the brain. The presented results open avenues for collaborative forms of AI as tools that naturally integrate with human cognition and allow collaboration directly from brain signals, especially in the settings of visual information generation. As a result, implicit EEG feedback displays potential for breakthrough applications that support image generation tasks, aggregating users' opinions when perceiving visual information in novel human-in-the-loop approaches.

ACKNOWLEDGMENT

Computing resources were provided by the Finnish Grid and Cloud Infrastructure (persistent id: urn:nbn:fi:research-infras-2016072533). The authors thank Jaakko Lehtinen, Tero Karras, Samuli Laine, and Timo Aila from NVIDIA for providing assistance and advice on GANs.

REFERENCES

[1] H. Ahmed, R. B. Wilbur, H. M. Bharadwaj, and J. M. Siskind, "Confounds in the data—Comments on decoding brain representations by multimodal learning of neural activity and visual features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 9217–9220, Dec. 2022, doi: 10.1109/TPAMI.2021.3121268.

[2] H. Ahmed, R. B. Wilbur, H. M. Bharadwaj, and J. M. Siskind, "Object classification from randomized EEG trials," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 3845–3854.

[3] S. F. Ahmed et al., "Deep learning modelling techniques: Current progress, applications, advantages, and challenges," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 13521–13617, Nov. 2023. [Online]. Available: https://doi.org/10.1007/s10462-023-10466-8

[4] H. M. Bharadwaj, R. B. Wilbur, and J. M. Siskind, "Still an ineffective method with supertrials/ERPs—Comments on 'decoding brain representations by multimodal learning of neural activity and visual features'," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 14052–14054, Nov. 2023.

[5] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of ERP components—A tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814–825, 2011.

[6] M. G. Bleichner and S. Debener, "Concealed, unobtrusive ear-centered EEG acquisition: CEEGrids for transparent EEG," *Front. Human Neurosci.*, vol. 11, p. 163, Apr. 2017.

[7] J.-H. Cho, J.-H. Jeong, and S.-W. Lee, "NeuroGrasp: Real-time EEG classification of high-level motor imagery tasks using a dual-stage deep learning framework," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13279–13292, Dec. 2022.

[8] E. Courchesne, S. A. Hillyard, and R. Galambos, "Stimulus novelty, task relevance and the visual evoked potential in man," *Electroencephalogr. Clin. Neurophysiol.*, vol. 39, no. 2, pp. 131–143, 1975.

[9] E. Cutrell and D. Tan, "BCI for passive input in HCI," in *Proc. CHI*, vol. 8, 2008, pp. 1–3.

[10] K. M. Davis, L. Kangassalo, M. Spapé, and T. Ruotsalo, "Brainsourcing: Crowdsourcing recognition tasks via collaborative brain-computer interfacing," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2020, pp. 1–14.

[11] E. Donchin and M. G. Coles, "Is the P300 component a manifestation of context updating?" *Behav. Brain Sci.*, vol. 11, no. 3, pp. 357–374, 1988.

[12] M. J. Eugster et al., "Predicting term-relevance from brain signals," in *Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2014, pp. 425–434.

[13] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2414–2423.

[14] P. Good, *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*, 2nd ed. New York, NY, USA: Springer, 2000, pp. 33–65.

[15] I. J. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[16] J. Grizou, I. Iturrate, L. Montesano, P.-Y. Oudeyer, and M. Lopes, "Calibration-free BCI based control," in *Proc. AAAI Conf. Artif. Intell.*, vol. 28, 2014, pp. 1–8.

[17] L. Guo, J. Liu, Y. Wang, Z. Luo, W. Wen, and H. Lu, "Sketch-based image retrieval using generative adversarial networks," in *Proc. 25th ACM Int. Conf. Multimedia*, 2017, pp. 1267–1268.

[18] A. G. Habashi, A. M. Azab, S. Eldawlatly, and G. M. Aly, "Generative adversarial networks in EEG analysis: An overview," *J. Neuroeng. Rehabil.*, vol. 20, no. 1, p. 40, 2023. [Online]. Available: https://doi.org/10.1186/s12984-023-01169-w

[19] G. Hajcak and D. Foti, "Significance?& Significance! Empirical, methodological, and theoretical connections between the late positive potential and P300 as neural responses to stimulus significance: An integrative review," *Psychophysiology*, vol. 57, no. 7, 2020, Art. no. e13570.

[20] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Human Mental Workload*, vol. 52, pp. 139–183, Jan. 1988.

[21] C. He et al., "Diversity and suitability of the state-of-the-art wearable and wireless EEG systems review," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 8, pp. 3830–3843, Aug. 2023.

[22] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, 2020, pp. 1–12.

[23] C. D. Hundhausen, D. Fairbrother, and M. Petre, "An empirical study of the 'prototype Walkthrough': A studio-based activity for HCI education," *ACM Trans. Comput.-Hum. Interact.*, vol. 19, no. 4, pp. 1–36, Dec. 2012.

[24] K. Davis and T. Ruotsalo, "Physiological data: Challenges for privacy and ethics," 2024, *arXiv:2405.15272*.

[25] B. Kaneshiro, M. Perreau Guimaraes, H.-S. Kim, A. M. Norcia, and P. Suppes, "A representational similarity analysis of the dynamics of object processing using single-trial EEG classification," *Plos ONE*, vol. 10, no. 8, pp. 1–27, Aug. 2015.

[26] L. Kangassalo, M. Spapé, and T. Ruotsalo, "Neuroadaptive modelling for generating images matching perceptual categories," *Sci. Rep.*, vol. 10, Sep. 2020, Art. no. 14719.

[27] A. Kapoor, P. Shenoy, and D. Tan, "Combining brain computer interfaces with vision for object categorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.

[28] A. Kapoor, D. Tan, P. Shenoy, and E. Horvitz, "Complementary computing for visual tasks: Meshing computer vision with human visual processing," in *Proc. IEEE Conf. Autom. Face Gesture Recognit.*, 2008, pp. 1–7.

[29] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. CVPR*, 2019, pp. 4401–4410.

[30] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. ICLR*, 2018, pp. 1–26.

[31] I. Kavasidis, S. Palazzo, C. Spampinato, D. Giordano, and M. Shah, "Brain2Image: Converting brain signals into images," in *Proc. 25th ACM Int. Conf. Multimedia*, 2017, pp. 1809–1817.

[32] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. 2nd Int. Conf. Learn. Represent.*, 2014, pp. 1–14.

[33] R. Li et al., "The perils and pitfalls of block design for EEG classification experiments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 316–333, Jan. 2021.

[34] G. Lindgaard, G. Fernandes, C. Dudek, and J. Brown, "Attention Web designers: You have 50 milliseconds to make a good first impression!" *Behav. Inf. Technol.*, vol. 25, no. 2, pp. 115–126, 2006.

[35] V. Liu and L. B. Chilton, "Design guidelines for prompt engineering text-to-image generative models," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2022, pp. 1–23.

[36] Y. Liu, Q. Li, Q. Deng, Z. Sun, and M.-H. Yang, "GAN-based facial attribute manipulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 12, pp. 14590–14610, Dec. 2023, doi: 10.1109/TPAMI.2023.3298868.

[37] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 3730–3738.

[38] Y. Lu, S. Wu, Y.-W. Tai, and C.-K. Tang, "Image generation from sketch constraint using contextual GAN," in *Proc. ECCV*, 2018, pp. 1–16.

[39] A. Luximon and R. S. Goonetilleke, "Continuous subjective workload assessment technique," in *Proc. First World Congr. Ergonom. Global Qual. Productiv.*, 1998, pp. 68–71.

[40] X. Ma, L. Yu, J. L. Forlizzi, and S. P. Dow, "Exiting the design studio: Leveraging online participants for early-stage design feedback," in *Proc. 18th ACM Conf. Comput. Support. Cooper. Work Soc. Comput.*, 2015, pp. 676–685.

[41] J. N. Mandrekar, "Receiver operating characteristic curve in diagnostic test assessment," *J. Thorac. Oncol.*, vol. 5, no. 9, pp. 1315–1316, 2010.

[42] C. D. Morgan, M. W. Geisler, J. W. Covington, J. Polich, and C. Murphy, "Olfactory P3 in young and older adults," *Psychophysiology*, vol. 36, no. 3, pp. 281–287, 1999.

[43] M. Ojala and G. C. Garriga, "Permutation tests for studying classifier performance," in *Proc. 9th IEEE Int. Conf. Data Min.*, 2009, pp. 908–913.

[44] S. Palazzo, C. Spampinato, I. Kavasidis, D. Giordano, J. Schmidt, and M. Shah, "Decoding brain representations by multimodal learning of neural activity and visual features," *IEEE Trans. Pattern Anal., Mach. Intell.*, vol. 43, no. 11, pp. 3833–3849, Nov. 2021, doi: 10.1109/TPAMI.2020.2995909.

[45] S. Pancholi, A. Giri, A. Jain, L. Kumar, and S. Roy, "Source aware deep learning framework for hand kinematic reconstruction using EEG signal," *IEEE Trans. Cybern.*, vol. 53, no. 7, pp. 4094–4106, Jul. 2023.

[46] "Artificial intelligence act: European parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European parliament and of the council on laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union Legislative Acts (COM(2021)0206–C9-0146/2021–2021/0106(COD))," Eur. Parliam., Strasbourg, France, document P9_TA(2024)0138, 2024.

[47] F. Pedregosa et al., "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, no. 85, pp. 2825–2830, 2011.

[48] J. Polich, "Updating P300: An integrative theory of P3a and P3b," *Clin. Neurophysiol.*, vol. 118, no. 10, pp. 2128–2148, 2007.

[49] A. Ramesh et al., "Zero-shot text-to-image generation," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 8821–8831.

[50] J. J. Rocchio, *Relevance Feedback in Information Retrieval*. Englewood, Cliffs, NJ, USA: Prentice-Hall, 1971.

[51] T. Ruotsalo, V. J. Traver, A. Kawala-Sterniuk, and L. A. Leiva, "Affective relevance," *IEEE Intell. Syst.*, early access, Apr. 19, 2024, doi: 10.1109/MIS.2024.3391508.

[52] T. Ruotsalo, K. Mäkelä, M. M. Spapé, and L. A. Leiva, "Feeling positive? Predicting emotional image similarity from brain signals," in *Proc. 31st ACM Int. Conf. Multimedia*, 2023, pp. 5870–5878. [Online]. Available: https://doi.org/10.1145/3581783.3613442

[53] B. Serim and G. Jacucci, "Explicating 'implicit interaction': An examination of the concept and challenges for research," in *Proc. 2019 CHI Conf. Human Factors Comput. Syst.*, 2019, pp. 1–16.

[54] F. X. Shen, S. M. Wolf, R. G. Gonzalez, and M. Garwood, "Ethical issues posed by field research using highly portable and cloud-enabled neuroimaging," *Neuron*, vol. 105, no. 5, pp. 771–775, 2020.

[55] P. Shenoy and D. S. Tan, "Human-aided computing: Utilizing implicit human processing to classify images," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2008, pp. 845–854.

[56] A. Shoshan, N. Bhonker, I. Kviatkovsky, and G. Medioni, "GAN-control: Explicitly controllable GANs," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 14083–14093.

[57] C. Spampinato, S. Palazzo, I. Kavasidis, D. Giordano, N. Souly, and M. Shah, "Deep learning human mind for automated visual classification," in *Proc. CVPR*, 2017, pp. 1–9.

[58] M. Spape, K. Davis, L. Kangassalo, N. Ravaja, Z. Sovijarvi-Spape, and T. Ruotsalo, "Brain-computer interface for generating personally attractive images," *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 637–649, Jan.–Mar. 2023.

[59] M. Spapé, R. Verdonschot, and H. V. Steenbergen, *The E-Primer: An Introduction to Creating Psychological Experiments in E-Prime*, 2nd ed. Zuid-Holland, The Netherlands: Leiden Univ., 2019.

[60] N. K. Squires, K. C. Squires, and S. A. Hillyard, "Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man," *Electroencephalogr. Clin. Neurophysiol.*, vol. 38, no. 4, pp. 387–401, 1975.
[61] G. Stefanics et al., "Cross-modal visual–auditory–somatosensory integration in a multimodal object recognition task in humans," *Int. Congr. Ser.*, vol. 1278, pp. 163–166, Mar. 2005.
[62] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," *Nature*, vol. 381, pp. 520–522, Jun. 1996.
[63] P. Tirupattur, Y. S. Rawat, C. Spampinato, and M. Shah, "ThoughtViz: Visualizing human thoughts using generative adversarial network," in *Proc. 26th ACM Int. Conf. Multimedia*, 2018, pp. 950–958.
[64] A. Todorov, *Face Value: The Irresistible Influence of First Impressions*. Princeton, NJ, USA: Princeton Univ., 2017.
[65] M. Tohidi, W. Buxton, R. Baecker, and A. Sellen, "Getting the right design and the design right," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2006, pp. 1243–1252.
[66] C. de la Torre-Ortiz, M. Spapé, and T. Ruotsalo, "The P3 indexes the distance between perceived and target image," *Psychophysiology*, vol. 60, no. 5, 2023, Art. no. e14225.
[67] C. de la Torre-Ortiz, M. M. Spapé, L. Kangassalo, and T. Ruotsalo, "Brain relevance feedback for interactive image generation," in *Proc. 33rd Annu. ACM Symp. User Interface Softw. Technol.*, 2020, pp. 1060–1070.
[68] A. Ukkonen, P. Joona, and T. Ruotsalo, "Generating images instead of retrieving them: Relevance feedback on generative adversarial networks," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2020, pp. 1329–1338.
[69] M. D. Vos, M. Kroesen, R. Emkes, and S. Debener, "P300 speller BCI with a mobile EEG system: Comparison to a traditional amplifier," *J. Neural Eng.*, vol. 11, no. 3, Apr. 2014, Art. no. 36008.
[70] J. Xu and C. Zheng, "Linear semantics in generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 9351–9360.
[71] T. O. Zander, L. R. Krol, N. P. Birbaumer, and K. Gramann, "Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity," *Proc. Nat. Acad. Sci.*, vol. 113, no. 52, pp. 14898–14903, 2016.
[72] D. Zhang, L. Yao, K. Chen, S. Wang, X. Chang, and Y. Liu, "Making sense of spatio-temporal preserving representations for EEG-based human intention recognition," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3033–3044, Jul. 2020.
[73] H. Zhang et al., "StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 5908–5916.
[74] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 839–847, Mar. 2019.
[75] P. Zhu, R. Abdal, Y. Qin, and P. Wonka, "SEAN: Image synthesis with semantic region-adaptive normalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 1–19.

Carlos de la Torre-Ortiz is currently pursuing the Ph.D. degree in computer science with the University of Helsinki, Helsinki, Finland.

His research interests lie in the fields of cognitive computing and machine learning.

Michiel M. Spapé received the Ph.D. degree from Leiden University, The Netherlands, in 2009.

He is an Associate Professor of Cognitive Neuroscience with the Centre for Cognitive and Brain Sciences, Institute for Collaborative Innovation, University of Macau, Macau, China. He focuses on emotion, perception/action, and neuroadaptive interaction.

Niklas Ravaja received the Ph.D. degree in psychology from the University of Helsinki, Helsinki, Finland, in 1996.

He is a Full Professor of eHealth and Well-Being with the University of Helsinki. He is an Expert on emotional and physiological processes during mediated social interaction.

Tuukka Ruotsalo received the Dr.Sc. (Tech) degree from Aalto University, Espoo, Finland, in 2010.

He is an Associate Professor with the University of Copenhagen, Copenhagen, Denmark, and LUT University, Lappeenranta, Finland. His research interests include machine learning, information retrieval, and cognitive and physiological computing.