Emotion-Guided Image and Text Generation: A Framework for Classifying EEG-Based Brain Signal Manifolds to Create Emotion-Sensitive Outputs

Linthe van Rooij
Media Technology
Leiden University
Leiden, The Netherlands
lintherooij@tudelft.nl

José Bençăo Fernandes
Mechanical Engineering
Delft University of Technology
Delft, The Netherlands
j.fernandes@student.tudelft.nl

Hidde Buiting
Biomedical engineering
Delft University of technology
Delft, The Netherlands
h.y.buiting@student.tudelft.nl

***Abstract*—In recent years, generative artificial intelligence has become more relevant in society. With this growing demand, prompt engineering for these artificial intelligence models is becoming more relevant. For some purposes in society, it can be useful to incorporate human emotion in the generated output. A way to include this is by measuring emotion with neuroimaging. Emotion can be spatially mapped using valence and arousal as axes, and therefore measuring these two parameters serves as a method to measure emotion. In this work, a framework is proposed that generates images and text guided by human emotions extracted from EEG signals using a convolutional neural network. In addition, a pilot study was conducted to evaluate the effectiveness of this framework by having participants estimate an emotion corresponding to an image. The pilot study showed that generative models are able to incorporate emotions. More research is required on the strategies to improve these accuracies.**

***Index Terms*—Artificial intelligence, EEG, emotional prompting, image generation.**

I. Introduction

In recent years, generative artificial intelligence (AI) has advanced to a stage where it shows impressive achievements in text-to-image and text-to-text creation [1, 2]. Generative models can now produce detailed and contextually relevant images and text based on a textual prompt. Due to these advancements, a challenge gaining relevance is prompt engineering. Prompts are instructions given to an AI model to enforce rules, automate processes, and ensure specific qualities and quantities of generated output. Prompts are also a form of programming that can customize the outputs and interactions with an AI model [3]. With prompt engineering, the details and context of created images are adjusted to fit many purposes.

However, prompt engineering does not yet consider human emotional data [4, 5]. Human emotion is important to be included in AI models because it can enhance user satisfaction and trust in the model [6]. Moreover, it is important because humans are not neutral themselves. People approach AI models with inherent biases, emotions, and subjective experiences that influence their interactions with technology. Expecting neutral outcomes from AI models, when the users themselves are not neutral, overlooks the complexity of human behaviour and decision-making.

Including emotion in prompting to generate images and text can help in the design of products and the way that these products are advertised, for example, it can be useful to create scenery linked to emotions to place the product in. It is useful for designers to create designs that are associated with certain emotions, allowing them to effectively address and engage with people's emotional responses. Various design models can be adopted for an efficient design process, such as the double diamond model, [7]. The generated images are tested by comparing them to an image generated without a specified emotion in the prompt.

It can be difficult to incorporate human emotion into prompting, a potential way to include it may be measuring emotions. This can be done with various neuroimaging modalities such as magnetoencephalography, functional magnetic resonance imaging or electroencephalography (EEG). Due to its high temporal resolution and easy applicability, EEG plays a major role in research on brain signals [8]. Identifying EEG-based discriminative features for visual categorization might provide meaningful insight about the human visual perception systems. This can lead to a new form of brain-based image labeling.

In this research, we propose a framework to generate images and text guided by human emotions extracted from EEG data. The goal of this framework is to learn a brain signal discriminative manifold of emotional categories by classifying EEG signals and to test whether this manifold can be used as input to generate emotion-sensitive outputs. Emotional prompting is central in this framework since it functions as a way to generate emotion-sensitive output based on human thought acquired through EEG.

We address the emotions used in the prompts with a classification model based on EEG data and translate this into an emotion that can be used in the prompt. For this translation, we adopt a spatial model that maps emotions described by Russell *et al.* [9]. In Russell's study, affective dimensions are studied such as pleasure, excitement, arousal, distress,

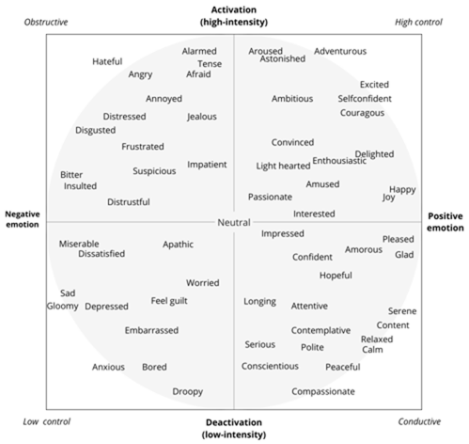


Fig. 1: Two-dimensional valence and arousal model. Figure adapted from Paltoglou and Thelwall [10].

displeasure, depression, sleepiness and relaxation. The authors found evidence that these affective dimensions are interrelated. The evidence suggests that these interrelationships can be represented by a spatial model in which affective concepts fall in a circle as shown in Fig. 1. With this spatial model, emotions can be determined with two parameters, valence and arousal.

Using the valence and arousal predictions from the classification model, we propose to enrich four neutral prompts in the stages of the double diamond design process for each of the predicted classes. These prompts will then be used as input for both OpenAI's GPT-4o model to generate emotion-guided text and Dall-E-3 to generate emotion-guided images.

Additionally, we conduct a small pilot study to evaluate the effectiveness of this framework. For this study, human evaluators will rate each generated image on emotions, using the valence and arousal scales [9]. With this pilot study we will provide meaningful insight into how effectively emotional input can generate emotion-guided output.

Taken together these steps in the process, our framework will provide insight into whether EEG can be a valuable asset for prompting to generate emotion-sensitive outputs. Therefore, we will have the following contributions. **1)** We propose a framework that can incorporate EEG data to effectively generate emotion-sensitive output. **2)** A classification model is used to translate EEG data into meaningful emotional values that can be used in prompting. **3)** A pilot study is done to demonstrate the framework's effectiveness.

II. Related work

A. Extracting emotions from EEG data

Electroencephalography (EEG) is a widely used noninvasive neuroimaging modality. Its main applications are in the assessment of cerebral function and it is the main source of signals for noninvasive brain-computer interfaces [11]. EEG is a recording of brain signals, also known as electric potentials. EEG signals contain several brain wave frequencies that correlate with specific mental and emotional states. These frequencies can be divided into five bands: delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–64 Hz). Each band has unique associations with neural activity, where delta waves are linked to deep sleep and theta to relaxation and creativity. Alpha waves are commonly observed during relaxed alertness, beta waves during active thinking and concentration, and gamma waves, during high-level cognitive processing and attention [12]. Accordingly, various emotions can be extracted from these frequency waves. Correlations between these waves and emotional states are described in literature [13–17]. In addition, research has been done on automated emotion recognition from EEG data using a multilayer perceptron neural network by Mrjit *et al.* [18].

B. Generating images and text based on EEG data

EEG data has been studied as input for generative AI in previous studies. For example, Kavasidis *et al.* [19] have proposed Brain2Image, a model to generate images using EEG data. The goal of these images was that they were semantically coherent with the visual stimuli evoking those brain responses. Something similar was done by Tirupattur *et al.* [20]. They have proposed ThoughtViz, a model that aims to visualize thoughts. A person is tasked to think of an object, an EEG is recorded and based on these signals, the model generates a visual reconstruction of such an object. Bai *et al.* [21] proposed DreamDiffusion, a model that uses EEG as input for an image generative model. It aims to reconstruct visual stimulus just like Brain2Image, and generally performs better.

III. Methodology

The proposed framework consists of using EEG data signals to first obtain a classification of valence and arousal and using that to generate text descriptions and images embedded with said classifications, as described by the schematic in Fig.2.

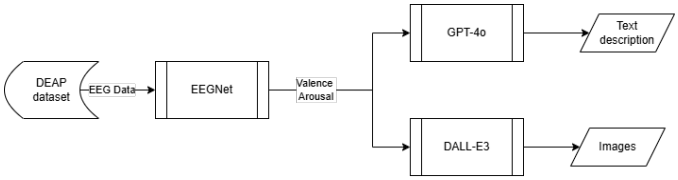


Fig. 2: Simplified project architecture

A. EEG as input for emotion classification

The dataset used in this study was the Database for Emotion Analysis using Physiological Signals (DEAP) dataset, this is a dataset designed for research in affective computing and emotion recognition [17]. The dataset contains EEG signals that were reshaped into a three-dimensional structure of shape (1280, 32, 8064), representing 1280 samples with 32 EEG channels and each one with 8064 time instances. The labels, corresponding to valence and arousal levels, were encoded into binary classes "positive" and "negative" valence and "high" and "low" arousal based on a threshold of 5, using an ordinal

encoder. The dataset was also split 80%-20%, respectively for training and testing.

The model architecture, based on EEGNet [22], is designed to capture temporal and spatial features of EEG data consisting of three main blocks: Temporal Convolution, Depthwise Convolution, and Separable Convolution layers as shown in Fig. 3. In the first block, a temporal convolution is used to identify key frequency patterns in the EEG signals. This is followed by a depthwise convolution, which learns spatial filters for each channel individually, helping the model capture distinct spatial information from different brain areas. The third block of EEGNet uses separable convolutions to extract temporal features specific to each frequency and then combines these features with a pointwise convolution, which merges all the information across channels.

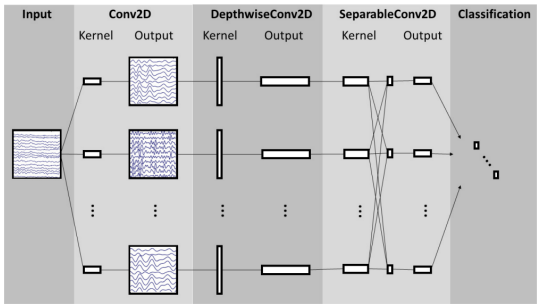


Fig. 3: EEGNet model architecture as shown in Lawhern *et al.* [22].

The exact structure can be found in Appendix A. The model training step was then performed for 50 epochs with a batch size of 64, utilizing the AdamW optimizer with CrossEntropy loss. A step-learning rate scheduler was also implemented with a step size of 5 and decay factor of 0.9, in order to improve convergence. Additionally, an early stopping criterion was also implemented, terminating training if no improvement in validation accuracy was observed for 10 consecutive epochs. The output of the model was reduced to only using valence and arousal labels each rated on a continuous scale from 1 to 9.

B. Image and text generation

After effectively predicting emotion classifications with the EEGNet model, we extract a valence/arousal pair for each of the classifications. We rescaled the output values of the model from $[1, 9]$ to $[-1, 1]$ and extracted the mean value for each classification as ground truth, thus the valence arousal pairs for each emotion as as follows: ”PHA”:[0.5,0.5];”PLA”:[-0.5,0.5];”NHA”:[0.5,-0.5];”NLA”:[-0.5,-0.5]. Incorporating the Double Diamond design process model as a guideline, we used four different prompts related to each of the design stages of the model [23].

The generative models to create the texts and images were both models from OpenAI to keep the training data source as identical as possible. The image generation used Dall-E-3, images were created with size 1024x1024 and standard quality. The text generation used GPT-4o. One additional prepromt was given to not use any markup or markdown in the output. For each stage, five images and text descriptions were generated based on the corresponding prompt. Four images and text descriptions were created with the added valence-arousal pairs mentioned previously, ”PHA,PLA,NHA,NLA”, and one image and text description without valence-arousal pair as a baseline image. This resulted in 20 images and 20 text descriptions in total.

IV. Experiments

A. Pilot study

The effectiveness of the framework was evaluated by a survey study. In total, 25 participants with ages ranging from 18-55 years (11 Female, 12 Male, and 2 Other) were eligible to participate in the survey since they already had prior experience with AI-generated images. Each participant was presented with the same 20 generated images across each stage and classification. After a practice trial, the stages were shown subsequently, while the order in which the classification was shown was randomized to rule out biases. The participants were instructed to rate each single presented image in the valence-arousal scale as shown in the task design in Fig 4. The blue dot represents the emotion that the participant decided was the best description and corresponds to an X and Y value (valence and arousal pair). Additionally, for one image in each stage, some explanation about the participant’s choice was asked.

B. Sentiment analysis

To further elaborate the results of the pilot study, a sentiment analysis was performed to analyze the generated textual descriptions. The generated text descriptions were analyzed using the Naive Bayes Analyzer from the Natural Language Processing Tool Kit. For each text description, this created a probability of negative or positive sentiment.

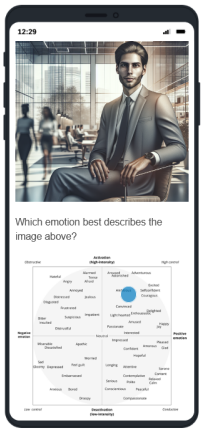


Fig. 4: Task design of pilot study on mobile phone as presented to the participants

V. Results

A. Dataset - Label frequency

In order to better understand the data at hand it is relevant to plot the distribution of labels in each one of the classification scenarios: only the valence, only the arousal, and both scales simultaneously. From both Table I and II it is clear that the positive valence and high arousal labels are more common than the remaining labels, being 54.7% more frequent than the second most common label in the combined labels scenario.

TABLE I: Label distribution for only arousal and only valence cases

Label count	Valence
Positive	724
Negative	556

Label count	Arousal
High	754
Low	526

TABLE II: Label distribution for combined labels case

Combined label count		Valence	
		Positive	Negative
Arousal	High	458	296
	Low	266	260

B. Classification model

Regarding the model's performance, 3 models were trained to classify: only the valence, only the arousal, and both scales simultaneously, allowing for comparison with similar models proposed in the literature. One of the similar setups is the one proposed by Chao, *et al.* [24]. Additionally, results for various models were compiled by S. Marjit, *et al.* [18] and are presented in Table III.

TABLE III: Model accuracy comparison

	Accuracy (%)		
Model/Classifier	**Valence**	**Arousal**	**Combined**
EEGNet	65.62	61.33	38.28
CapsNet	66.73	68.28	N/A
SVM (3rd order polynomial kernel)	60.16	66.80	41.66
SVM (RBF kernel)	56.64	66.64	40.63
Random Forest	63.67	63.67	43.40
XgBoost	64.84	68.98	44.23
k-Nearest Neighbors	64.45	62.89	44.28
Linear Regression	67.20	61.72	42.19
Dummy Classifier	50	50	25

As shown in Table III, the accuracy of the models for the individual valence and arousal classification is similar and overall better than the dummy classifier, regarding the combined labels, the used model presents a lower accuracy than the other models, but still better than the random chance accuracy, given by the dummy classifier. It is also worth mentioning that no data was available for the CapsNet model in the combined label scenario.

Plotting the confusion matrices it is possible to attain deeper insights into how the model performs for each distinct label, and which labels are more often misclassified. From visual inspection of Fig. 5 it is clear that generally, the model classifies data as positive high arousal more often than others.

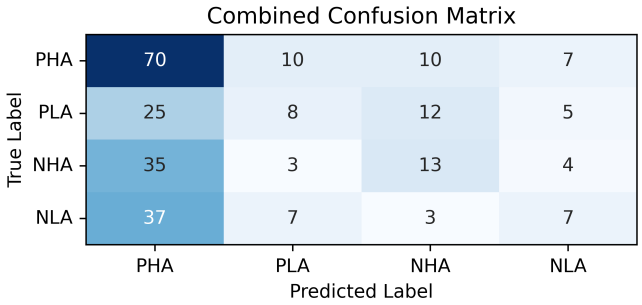


Fig. 5: Confusion matrix for combined valence and arousal case

C. Generative model

The 20 generated images obtained from the generative model are listed in Appendix B.

D. Pilot study

In Fig. 8a the original data from the survey is shown without baseline correction for all 20 images across 25 participants. Each color corresponds to one of the classes and crosses correspond to each ground truth valence-arousal pair. Here, the data shows a bias towards a positive valence. However, after subtracting each data point with a within-subject baseline correction for each stage, shown in Fig. 8b, the data is more distributed across the valence-arousal space. We added a limit of -1 and 1 for each baseline correction to prevent exceeding our original scale. In Fig. 6 the data is separated into their own class. From this data, no clear clusters can be seen around the ground truth points. Furthermore, euclidean distance measures between the average of the participant's scores for each class are between 0.78 and 0.9 as shown in table IV. If we relate this to the range of the total valence and arousal scale, the Euclidean distance scores range from 39% to 45% of the total range.

Classifying the evaluated values into the classes from the classification model and comparing them to the labels of the expected values that we put in the prompts will give the confusion matrix as shown in Fig. 7. This is returning an accuracy of 0.22.

Furthermore, all generated text across all classifications was labeled as positive descriptions by the sentiment analysis tool.

TABLE IV: Results based on Euclidean Distance

Category	**Average Distance**	**Percentage**
NHA	0.7750	38.75%
NLA	0.9073	45.36%
PHA	0.8914	44.57%
PLA	0.8725	43.63%

VI. Discussion

A. Dataset and classification accuracy

Regarding the accuracies achieved by the implemented model, these are on par with models proposed in the literature

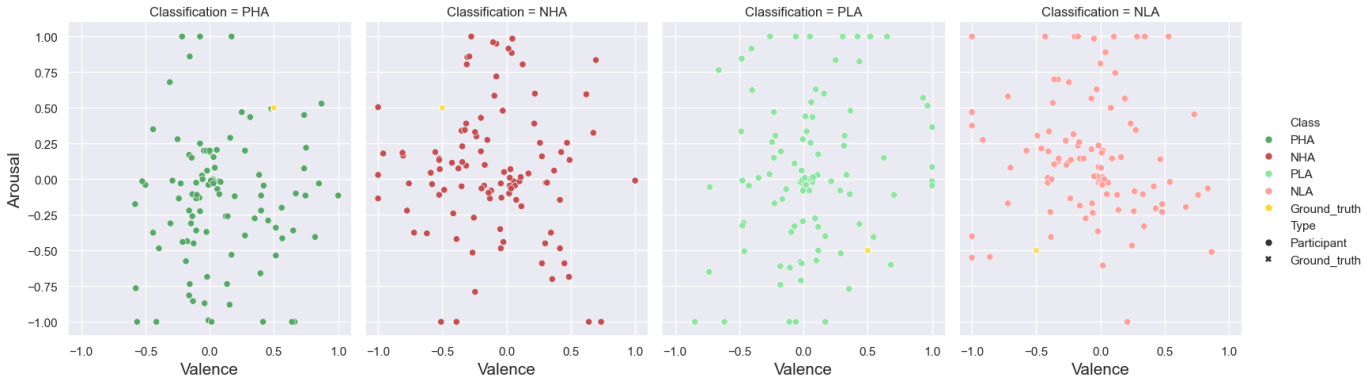


Fig. 6: Normalized results for each classification

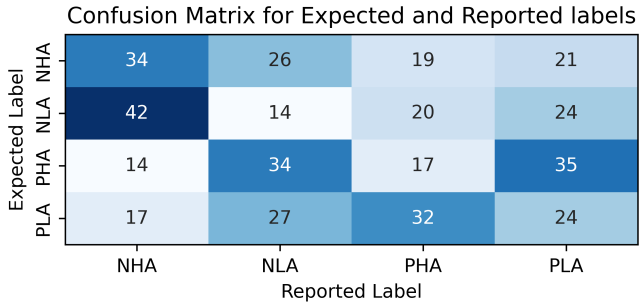
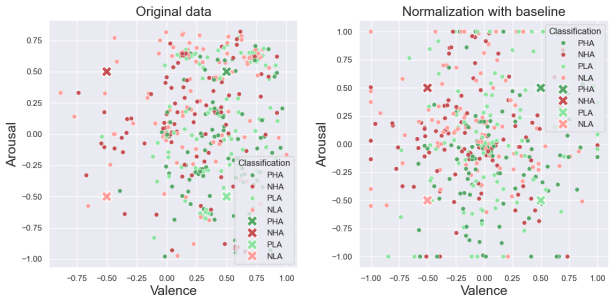


Fig. 7: Confusion Matrix of the actual valence-arousal pairs and the evaluated valence-arousal pairs by the participants



(a) Original Image (b) Normalized Image

Fig. 8: Comparison of Normalized and Original data

that display similar complexity and use the same dataset. Some more intricate models that achieve better accuracies have been proposed, such as the one by S. Marjit *et al.* in [18], but at the expense of high complexity, low explainability and high computational cost. Additionally, given the limited access to high-performance computational hardware, the number of training iterations was limited, which constrained the performance achievable by the implemented model. About the attained confusion matrices, these are a clear reflection of the bias present in the dataset. Given that the DEAP dataset has more positive and high arousal emotions, as shown in Table I and II, the model will be trained to better detect specifically these emotions, and in uncertain scenarios, it will lean towards these labels. In order to improve on the presented results, a more comprehensive dataset could be used and the complexity of the model could be increased.

B. Generative model

As it can be seen in the results shown in Section V-C, the generative models were able to embed emotions into the images, but not in an accurate way. This might be a result of these models not having been trained to keep emotions in mind, which is a complex task given the inherent subjectivity associated with trying to predict the emotional response that a piece of information will cause in different people. Additionally, more often than not, these generative models have put in place mechanisms that prevent the generation of disturbing images or images associated with possible negative emotions, which explains the clear trend towards the right-half plane shown in Figure 8a.

C. Societal impact

A model that is able to generate text and images embodying human emotion can be of use in many applications in society. When people in the future may have access to brain computer interfaces, it may be of help if the computer is able to understand and respond to the user's emotion. For example, when the model recognizes that the user is distressed, a visual stimulus can be shown with a relaxing purpose. Therefore, it can be of help for people struggling with mental health issues. Another one of these applications is in the creative industry, the emotional state of the designer can be translated into a visual representation of the product. This can enhance designing processes by providing more inspiration, especially when a generated image needs to evoke a certain feeling.

D. Ethical considerations

Ethical considerations play a role in implementing frameworks such as suggested by this paper. An ethical consideration that has to be made involves the security and confidentiality of the EEG data. EEG data needs to be protected against potential leaks, because this may contain intimate emotional or mental information. Another issue to consider is that this

technique may be used for the wrong reasons or output may be wrongly interpreted. This could especially be of harm when it is used in a legal or medical setting. In addition, people contributing EEG data must be well informed on how their EEG data is used and what could be the potential outcomes of the model. If people who contribute EEG data are not completely aware of the way that their data is used or of the potential implications that the output may have, their autonomy is harmed.

VII. Conclusion

This research proposes a framework to incorporate EEG data into prompt engineering to generate emotion-sensitive outputs. Translating EEG data into meaningful emotional values can be used as a method to incorporate human emotions into prompting to generate emotion-guided output. Moreover, with the pilot study, we showed that generative models are capable of incorporating emotions into their output. However, more research needs to be done to conclude on the accuracy of these models and the strategies to improve them.

References

[1] Stefan Feuerriegel et al. "Generative ai". In: *Business & Information Systems Engineering* 66.1 (2024), pp. 111–126.

[2] Roberto Gozalo-Brizuela and Eduardo C Garrido-Merchan. "ChatGPT is not all you need. A State of the Art Review of large Generative AI models". In: *arXiv preprint arXiv:2301.04655* (2023).

[3] Jules White et al. "A prompt pattern catalog to enhance prompt engineering with chatgpt". In: *arXiv preprint arXiv:2302.11382* (2023).

[4] Louie Giray. "Prompt engineering with ChatGPT: a guide for academic writers". In: *Annals of biomedical engineering* 51.12 (2023), pp. 2629–2633.

[5] Ggaliwango Marvin et al. "Prompt engineering in large language models". In: *International conference on data intelligence and cognitive informatics*. Springer. 2023, pp. 387–402.

[6] Endang Wahyu Pamungkas. "Emotionally-aware chatbots: A survey". In: *arXiv preprint arXiv:1906.09774* (2019).

[7] Design Council. "The 'double diamond' design process model". In: *Design Council* (2005). URL: https://www.designcouncil.org.uk/news-opinion/design-process-what-double-diamond.

[8] Gernot R Müller-Putz. "Electroencephalography". In: *Handbook of clinical neurology* 168 (2020), pp. 249–262.

[9] James A Russell. "A circumplex model of affect." In: *Journal of personality and social psychology* 39.6 (1980), p. 1161.

[10] Georgios Paltoglou and Michael Thelwall. "Seeing Stars of Valence and Arousal in Blog Posts". In: *IEEE Transactions on Affective Computing* 4.1 (2013), pp. 116–123. DOI: 10.1109/T-AFFC.2012.36.

[11] CD Binnie and PF Prior. "Electroencephalography." In: *Journal of Neurology, Neurosurgery & Psychiatry* 57.11 (1994), pp. 1308–1319.

[12] Ernst Niedermeyer and Fernando Lopes da Silva. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 2004.

[13] Daniela Sammler et al. "Music and emotion: electrophysiological correlates of the processing of pleasant and unpleasant music". In: *Psychophysiology* 44.2 (2007), pp. 293–304.

[14] Louis A Schmidt and Laurel J Trainor. "Frontal brain electrical activity (EEG) distinguishes valence and intensity of musical emotions". In: *Cognition & Emotion* 15.4 (2001), pp. 487–500.

[15] Ian Daly et al. "Neural correlates of emotional responses to music: an EEG study". In: *Neuroscience letters* 573 (2014), pp. 52–57.

[16] LI Aftanas et al. "Neurophysiological correlates of induced discrete emotions in humans: an individually oriented analysis". In: *Neuroscience and Behavioral Physiology* 36 (2006), pp. 119–130.

[17] Sander Koelstra et al. "DEAP: A database for emotion analysis; using physiological signals". In: *IEEE Transactions on Affective Computing* 3.1 (2012), pp. 18–31.

[18] Shyam Marjit, Upasana Talukdar, and Shyamanta M Hazarika. "EEG-Based Emotion Recognition Using Genetic Algorithm Optimized Multi-Layer Perceptron". In: *2021 International Symposium of Asian Control Association on Intelligent Robotics and Industrial Automation (IRIA)*. 2021, pp. 304–309. DOI: 10.1109/IRIA53009.2021.9588702.

[19] Isaak Kavasidis et al. "Brain2image: Converting brain signals into images". In: *Proceedings of the 25th ACM international conference on Multimedia*. 2017, pp. 1809–1817.

[20] Praveen Tirupattur et al. "Thoughtviz: Visualizing human thoughts using generative adversarial network". In: *Proceedings of the 26th ACM international conference on Multimedia*. 2018, pp. 950–958.

[21] Yunpeng Bai et al. "Dreamdiffusion: Generating high-quality images from brain eeg signals". In: *arXiv preprint arXiv:2306.16934* (2023).

[22] Vernon J Lawhern et al. "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces". In: *Journal of neural engineering* 15.5 (2018), p. 056013.

[23] Heng Gu. *Double Diamond Prompts*. 2024. URL: https://arteliers.notion.site/Double-Diamond-Prompts-13349f16dd2980c5bd26fd1f0997605f.

[24] Hao Chao et al. "Emotion Recognition from Multiband EEG Signals Using CapsNet". In: *Sensors* 19.9 (2019). ISSN: 1424-8220. DOI: 10.3390/s19092212. URL: https://www.mdpi.com/1424-8220/19/9/2212.