Evaluating Creative Expression through Integrated Image and Text Embeddings

Amaia Pikatza-Huerga[1][0009−0003−9080−6242], Pablo Matanzas de Luis[1][0009−0009−8897−5796], Miguel Fernandez-de-Retana Uribe[1][0009−0002−0883−1303], Javier Peña Lasa[2][0000−0002−0041−7020], Unai Zulaika[1][0000−0002−7366−9579], and Aitor Almeida[1][0000−0002−1585−4717]

[1] Faculty of Engineering, University of Deusto, Unibertsitate Etorb., 24, Bilbo, Spain
[2] Faculty of Health Science, University of Deusto, Unibertsitate Etorb., 24, Bilbo, Spain {a.pikatza, javier.pena, unai.zulaika, aitor.almeida}@deusto.es

Abstract. This study extends previous multimodal creativity assessment by employing CLIP (Contrastive Language–Image Pre-training) to generate unified embeddings of drawings and their accompanying titles. A dataset of 486 sketches—produced by 53 participants before and after intracranial stimulation—was annotated for originality (O), flexibility (FLE), elaboration (E), and title creativity (T) by expert psychologists. Eight classifiers and four regressors were evaluated on each target, comparing raw versus PCA-reduced CLIP features and applying SMOTE to balance the binary tasks (O, T). A Random Forest trained on raw CLIP embeddings achieved the highest AUC for originality (0.919), while an MLP regressor marginally improved elaboration R^2 (0.489) over image-only baselines. Text-based models remained strongest for title creativity, yet CLIP-based fusion delivered robust performance across all four dimensions. An accompanying Streamlit application allows users to complete a drawing in-browser and receive instantaneous creativity scores, demonstrating both technical advancement and practical applicability for educational and clinical settings.

Keywords: Machine Learning · Creativity Assessment · Artistic Expression · Text and Image Analysis

1 Introduction

The assessment of creativity has undergone significant advances through the integration of artificial intelligence (AI) and machine learning (ML), enabling more objective, scalable, and nuanced evaluation methods. Traditionally, creative tasks such as the Alternate Uses Task (AUT) or drawing-based assessments like the TCT-DP and Draw-A-Person Test have relied on human judgment, raising concerns about consistency, efficiency, and subjectivity [2,4,23].

With AI, it is now possible to quantify key dimensions of creativity—such as originality, flexibility, and elaboration—with greater precision. Tools like SemDis have automated the scoring of verbal tasks, while convolutional neural networks

(CNNs) have successfully replicated expert evaluations in drawing-based assessments [9]. These methods reduce both the time required and inter-rater variability.

More sophisticated models, such as the *DeepCreativity* framework [12], use deep learning techniques to assess creative output without predefined attributes. In addition, multimodal approaches that combine visual and textual data—such as drawing titles and descriptions—capture subtleties that single-modality models may miss, particularly in evaluating originality and flexibility [1,3,28].

AI has also contributed to a deeper understanding of the creative process. Recent studies have linked analyses of functional connectomes to traits like psychological resilience, showing the potential of AI to uncover relationships between creative output and underlying cognitive characteristics [25]. This highlights AI's role not only as an evaluative tool but also as a means to enhance theoretical understanding of creativity.

In emotional and therapeutic contexts, AI has shown promise in analyzing affective content in drawings through sentiment analysis techniques [10,16,21]. These tools help clinicians monitor emotional and cognitive changes over time and detect deviations from normative profiles [18,22,29,11,24]. Furthermore, current models offer insights into cognitive styles and personality traits through drawing-based tasks [6,8,13], reinforcing AI's potential for comprehensive creativity assessment. Building on this progress, the present work introduces a multimodal model that integrates visual and textual data for a more complete and human-aligned evaluation of creativity [14].

2 Methodology

2.1 Data

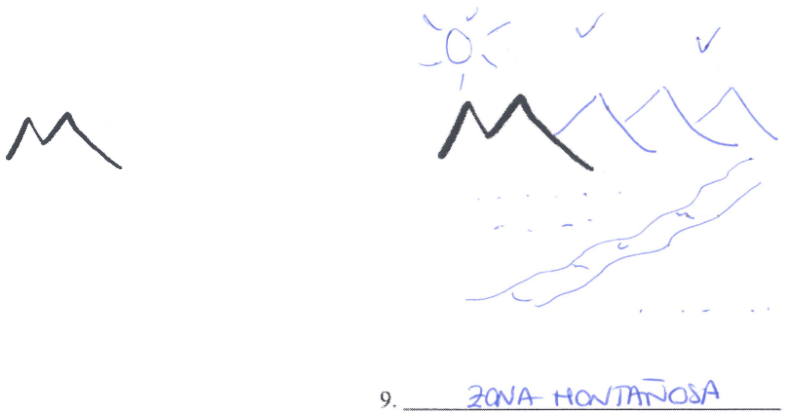
We use the multimodal dataset introduced in Pikatza-Huerga et al. (2025), consisting of 486 samples of scanned drawings with accompanying Spanish titles [19]. These drawings were collected in a study of intracranial stimulation effects on creativity: 53 participants produced one drawing before stimulation (pre-stimulation phase) and one drawing after stimulation (post-stimulation phase). For each drawing, participants also provided a short title or description reflecting their interpretation of the image.

The primary aim of the original study was to assess whether intracranial stimulation influences drawing originality. The resulting dataset includes both pre- and post-stimulation drawings and titles, serving as the basis for all our experiments. Each sample comprises:

- **IMAGE:** A 224×224 px scanned drawing, normalized to $[0, 1]$.
- **TEXT:** The Spanish title supplied by the participant, lowercased, punctuation-stripped, and tokenized.
- **O (Originality):** Binary label (0 = not original, 1 = original) assigned by expert psychologists.

- **FLE (Flexibility):** Integer category count of distinct themes (e.g. people, landscapes), normalized to reflect thematic flexibility.
- **E (Elaboration):** Numeric score indicating the level of detail and complexity, as rated by experts.
- **T (Title Creativity):** Binary label (0 = not creative, 1 = creative) for the title, assigned by experts.

Figure 1 shows an example of a drawing line that participants were given to complete (Figure 1a) and the completed drawing (Figure 1b) with the title given.



(a) Base canvas for drawing (b) Completed drawing

Fig. 1: Scanned drawings

2.2 Participants

Fifty-three participants (49.1% male, 50.9% female; ages 10–60) provided drawings before and after intracranial stimulation. Demographics (mother tongue, education, sleep hours, stimulant/tobacco use) were recorded as in previous research [19].

2.3 Original Multimodal Models

In previous research, visual embeddings were obtained via CNN encoders—ResNet50, InceptionV3, EfficientNetB0, and Xception—and textual embeddings via BETO, FastText, or a Keras Embedding layer. These embeddings were concatenated and used to train classical machine-learning models: classifiers for originality (O), flexibility (FLE), and title creativity (T), as well as regressors

for elaboration (E). Performance was evaluated using ROC AUC, accuracy, recall, precision, specificity, and F1 score for classification tasks, and MAE, RMSE, and R^2 for regression tasks [19].

Building on this foundation, the original multimodal methodology included:

1. **Single-modality baselines:** Separate training with image-only and text-only embeddings to assess each modality's independent predictive power.
2. **Feature fusion:** Concatenation of the penultimate-layer embeddings from both modalities into a joint representation before prediction.
3. **Hyperparameter optimization:** Application of grid search (for smaller models) or randomized search (for larger parameter spaces) with k-fold cross-validation to tune key parameters such as tree depth, number of neurons, regularization strength, and learning rates.

Image-Based Models

- **ResNet50** [15]: A 50-layer deep residual network mitigating vanishing gradients via identity skip connections.
- **InceptionV3** [26]: Employs factorized convolutions and dimensionality reductions to improve computational efficiency.
- **EfficientNetB0** [27]: Utilizes compound scaling to balance network depth, width, and resolution for optimal accuracy-efficiency trade-offs.
- **Xception** [7]: Leverages depthwise separable convolutions for parameter-efficient feature extraction and improved classification performance.

Text Embedding Models

- **BETO** [5]: A Spanish-language BERT model providing contextualized word embeddings tailored to Spanish text.
- **FastText** [17]: Generates subword-aware embeddings that capture morphological information and handle out-of-vocabulary tokens.
- **Keras Embedding layer**: Learns dense word vectors end-to-end within the classification pipeline, offering a lightweight and efficient text representation.

This diverse selection of encoders ensures a comprehensive evaluation of how visual and textual modalities contribute—independently and jointly—to automated creativity assessment.

2.4 New CLIP-Based Experiment

To enhance joint image–text representation, we employ OpenAI's CLIP (ViT-B/32) [20] and systematically evaluate both PCA vs. no–PCA and SMOTE-balancing only on the binary tasks (O and T), while testing all classifiers/regressors on every target:

1. **Embedding extraction:**

   - Preprocess images (224×224 px, normalized) and titles (lowercased, tokenized).
   - Pass through CLIP's vision and text encoders to obtain 512-D vectors.
   - L_2–normalize both vectors.
2. **Feature fusion:**
   - Concatenate image and text vectors into a 1,024-D feature.
   - Optionally apply PCA to reduce to 256 dimensions (95% explained variance) via `PCA(n_components=256)` or skip PCA (leaving 1,024-D).
3. **Modeling and balancing:**
   - *Originality (O)* and *Title originality (T)* (binary):
     - Apply SMOTE to training folds to correct class imbalance.
     - Evaluate all classifiers (MLP, LogisticRegression, RandomForest, SVC, KNN, GaussianNB, DecisionTree, XGB) with and without PCA.
   - *Flexibility (FLE)* (multiclass):
     - No SMOTE; filter out classes with < 2 samples.
     - Evaluate all classifiers with and without PCA; optimize macro ROC AUC.
   - *Elaboration (E)* (regression):
     - No SMOTE.
     - Test all regressors (MLPRegressor, RandomForestRegressor, LinearRegression, XGBRegressor) with and without PCA; optimize negative MSE.
4. **Hyperparameter tuning:** For each model, perform `RandomizedSearchCV` (10 iterations, 3–fold CV) on either ROC AUC (classification) or negative MSE (regression).

2.5 Training and Evaluation

For each target dimension, data are split 80/20 with stratification to prevent participant overlap using `train_test_split`. Classification metrics include ROC AUC, accuracy, precision, recall, specificity (binary only), and F1 score; regression metrics include MSE, RMSE, and R^2. Results are exported to CSV for comparison against previous benchmarks.

2.6 Web Application

To demonstrate the practical applicability of the proposed models, an interactive web application was developed using Streamlit. The application enables users to complete a predefined base drawing on a digital canvas and submit a corresponding title. Upon submission, the drawing-title pair is evaluated in real time across four creativity dimensions: elaboration (E), flexibility (FLE), originality (O), and title creativity (T).

The application integrates several pretrained and fine-tuned models used in the study. Elaboration is predicted using a convolutional neural network trained on expert-labeled data. Flexibility scores are derived from FastText embeddings processed by a multiclass classifier. Title creativity is assessed using a Spanish

BERT model (BETO), while a CLIP-based pipeline combining image and text embeddings is used to compute the overall originality score. The CLIP embeddings are passed to a Random Forest classifier trained to distinguish original versus non-original submissions.

To promote user engagement and creative diversity, the system randomly selects a base drawing from a pool of initial sketches each time the user initiates a new session via a "Try Again" button. However, the drawing remains fixed during a session to preserve evaluation consistency. The final evaluation output includes a composite visualization of the user's completed drawing and the corresponding creativity metrics.

3 Results

This section compares the performance of models trained in the new CLIP-based experiment with those from previous research [19], across all four creativity-related prediction tasks: originality (O), elaboration (E), flexibility (FLE), and title creativity (T). For each task, we report the best-performing models in terms of ROC AUC for classification and R^2 for regression.

3.1 Predicting Originality (O)

In previous work, the best AUC was achieved with a combination of FastText and InceptionV3 (AUC = 0.85). In our new experiments, the Random Forest classifier trained on CLIP embeddings with no PCA outperformed all previous models, achieving an AUC of 0.92 and an accuracy of 0.83.

3.2 Predicting Elaboration (E)

In previous work, InceptionV3 achieved the best regression performance ($R^2 = 0.48$). In the new CLIP-based experiment, the MLPRegressor performed best with $R^2 = 0.49$, showing a slight improvement.

3.3 Predicting Flexibility (FLE)

The highest AUC in the original study was obtained using FastText alone (AUC = 0.91). In the new experiment, the best AUC was 0.87, obtained by an MLP classifier on CLIP embeddings. Although this represents a small drop in AUC, it came with improved balance across precision, recall, and F1.

3.4 Predicting Title Creativity (T)

As expected, text-only models performed best in predicting T. In previous research, BETO achieved an AUC of 0.91. In our new experiments, the best AUC was 0.84, obtained using a Random Forest classifier with CLIP text embeddings and PCA. Though slightly lower than BETO, the performance remains competitive.

Table 1: Performance for Originality (O)

Classifier	**AUC**	**Accuracy**	**Precision**	**Recall**	**F1**	**Specificity**
RandomForest	0.920	0.827	0.833	0.827	0.822	0.931
SVC + PCA	0.920	0.827	0.826	0.827	0.826	0.862
SVC	0.903	0.827	0.827	0.827	0.824	0.897
LogisticRegression + PCA	0.889	0.796	0.801	0.796	0.797	0.793
LogisticRegression	0.888	0.796	0.798	0.796	0.797	0.810
MLPClassifier + PCA	0.879	0.816	0.818	0.816	0.817	0.828
RandomForest + PCA	0.862	0.776	0.774	0.776	0.772	0.862
MLPClassifier	0.861	0.806	0.807	0.806	0.807	0.828
KNeighborsClassifier	0.858	0.745	0.775	0.745	0.747	0.672
KNeighborsClassifier + PCA	0.853	0.745	0.752	0.745	0.747	0.741
GaussianNB + PCA	0.844	0.765	0.764	0.765	0.765	0.810
XGBClassifier	0.855	0.735	0.732	0.735	0.731	0.828
XGBClassifier + PCA	0.810	0.735	0.735	0.735	0.735	0.776
GaussianNB	0.793	0.735	0.740	0.735	0.736	0.741
DecisionTreeClassifier + PCA	0.686	0.684	0.696	0.684	0.686	0.672
DecisionTreeClassifier	0.602	0.653	0.647	0.653	0.648	0.759
Baseline (prev.)	0.850	0.800	0.420	0.710	0.530	0.890

Table 2: Performance for Elaboration (E)

Regressor	**MSE**	**RMSE**	$\boldsymbol{R^2}$
MLPRegressor	2.102	1.450	0.489
RandomForestRegressor	2.393	1.547	0.418
XGBRegressor	2.495	1.579	0.393
LinearRegression	6.162	2.482	-0.498
Baseline (prev.)	2.820	1.300	0.480

3.5 Web application

The web application was evaluated in terms of its ability to replicate the scoring patterns observed in offline model testing. The system produced reliable and interpretable scores aligned with expert-labeled ground truth across all four dimensions. Visual elaboration scores increased proportionally with the amount of detail added to the base sketch. BETO-based evaluation effectively distinguished generic from inventive titles, while FastText-based classification provided coherent thematic interpretations for flexibility.

The CLIP-based originality classifier successfully identified novel and unexpected image-text combinations, with output probabilities offering fine-grained insight into the originality dimension. Randomized base drawings led to a wide range of user-generated responses, illustrating the robustness and generalizability of the evaluation framework.

Table 3: Performance for Flexibility (FLE)

Classifier	**AUC**	**Accuracy**	**Precision**	**Recall**	**F1**
MLPClassifier	0.874	0.558	0.551	0.558	0.537
MLPClassifier + PCA	0.857	0.568	0.541	0.568	0.532
RandomForestClassifier	0.842	0.537	0.552	0.537	0.487
RandomForestClassifier + PCA	0.837	0.568	0.570	0.568	0.515
LogisticRegression	0.834	0.495	0.509	0.495	0.430
LogisticRegression + PCA	0.833	0.495	0.513	0.495	0.433
GaussianNB + PCA	0.812	0.526	0.542	0.526	0.496
GaussianNB	0.791	0.484	0.507	0.484	0.441
SVC + PCA	0.768	0.526	0.512	0.526	0.468
SVC	0.762	0.526	0.512	0.526	0.468
KNeighborsClassifier + PCA	0.728	0.463	0.507	0.463	0.427
KNeighborsClassifier	0.719	0.453	0.499	0.453	0.413
Baseline (prev.)	0.910	0.560	0.800	0.370	0.660

This implementation confirms the potential of integrated, multimodal models for real-time creativity assessment. The application serves as a proof-of-concept for deploying machine learning-based evaluators in educational and clinical settings, enabling scalable and standardized creativity feedback.

To facilitate reproducibility and further development by the research community, the source code of the application is available on GitHub.

4 Discussion

The results demonstrate that leveraging CLIP embeddings for multimodal fusion yields notable improvements in certain creativity dimensions, while maintaining competitive performance in others compared to previous image-only and text-only baselines.

In the case of originality (O), the Random Forest classifier trained on raw CLIP embeddings achieved the highest AUC of 0.919, substantially exceeding the best baseline AUC of 0.850 reported for the FastText + InceptionV3 combination [19]. Other CLIP-based classifiers, such as SVC and MLP, also consistently outperformed prior benchmarks, with specificity values reaching up to 0.931 and F1 scores as high as 0.826. These results suggest that the joint image–text representation offered by CLIP is highly effective in capturing the originality dimension of creative work.

For elaboration (E), regression results showed that the MLP model trained on CLIP embeddings achieved an R^2 score of 0.489 and an RMSE of 1.450. While the R^2 represents a modest improvement over the best image-only baseline ($R^2 = 0.480$), the RMSE was slightly higher than that of the InceptionV3-based model (RMSE = 1.300). This indicates that although CLIP fusion improves explained variance, it may introduce additional variability that slightly affects error

Table 4: Performance for Title Creativity (T)

Classifier	**AUC**	**Accuracy**	**Precision**	**Recall**	**F1**	**Specificity**
RandomForestClassifier + PCA	0.838	0.786	0.794	0.786	0.788	0.787
RandomForestClassifier	0.830	0.684	0.675	0.684	0.676	0.803
XGBClassifier + PCA	0.829	0.704	0.696	0.704	0.694	0.836
SVC + PCA	0.799	0.765	0.762	0.765	0.761	0.853
LogisticRegression	0.786	0.735	0.735	0.735	0.735	0.787
LogisticRegression + PCA	0.785	0.714	0.712	0.714	0.713	0.787
MLPClassifier + PCA	0.786	0.776	0.773	0.776	0.771	0.869
MLPClassifier	0.785	0.786	0.784	0.786	0.780	0.885
XGBClassifier	0.807	0.684	0.675	0.684	0.676	0.803
DecisionTreeClassifier	0.715	0.735	0.731	0.735	0.731	0.820
KNeighborsClassifier + PCA	0.748	0.694	0.706	0.694	0.697	0.705
KNeighborsClassifier	0.712	0.673	0.727	0.673	0.677	0.590
GaussianNB	0.741	0.653	0.653	0.653	0.653	0.721
GaussianNB + PCA	0.630	0.592	0.601	0.592	0.595	0.639
Baseline (prev.)	0.910	0.900	1.000	0.800	0.900	0.740

magnitude. Nonetheless, the model remains a strong alternative for elaboration prediction.

In terms of flexibility (FLE), the MLP classifier using CLIP embeddings obtained an AUC of 0.874, compared to the 0.910 achieved by the FastText-only model in previous work. Although the CLIP-based models did not surpass the text-only baseline in AUC, they maintained competitive accuracy (0.558) and F1 score (0.537), showing the capacity of the fused embeddings to capture thematic variation without relying solely on textual cues.

Regarding title creativity (T), title classification remained a task where text-specific models excelled. The best BETO-based classifier achieved an AUC of 0.910 and F1 score of 0.900, whereas the top CLIP-based model (Random Forest with PCA) reached an AUC of 0.838 and F1 of 0.788. This suggests that, for short textual content like titles, specialized language models continue to provide the most accurate predictions, though visual context still contributes meaningful information.

These findings validate CLIP as a powerful tool for creativity assessment, particularly in tasks requiring integration of visual and textual semantics. The observed improvements in originality and elaboration predictions highlight CLIP's strength in capturing abstract creative qualities. While text-only models retain an edge in title-related tasks, CLIP provides a unified framework suitable for general multimodal prediction. The successful integration into an interactive Streamlit application further demonstrates its practical utility for real-time creativity evaluation in both educational and clinical settings.

4.1 Limitations

While the proposed CLIP-based framework shows promising results, several limitations should be noted.

First, the dataset and language models are based entirely on Spanish-language text. Although BETO and FastText are trained specifically for Spanish, CLIP was pretrained primarily on English data. This mismatch may affect the alignment between image and text modalities in Spanish and could explain the relatively lower performance of CLIP-based models in title creativity tasks.

Second, model interpretability remains limited. Deep learning architectures such as CLIP and MLPs are inherently difficult to interpret, and the absence of attention maps or visual explanations restricts the ability to understand which specific features influence the predictions.

Third, the dataset size is modest—486 examples from 53 participants—and may not capture the full variability of drawing styles, cultural contexts, or creative expressions. Generalization to broader populations remains an open question, particularly for fine-grained creativity assessments.

Lastly, while the Streamlit application provides real-time evaluation, the initial loading time for large models like BETO and CLIP can be substantial. This may pose limitations for low-resource or mobile deployments without further optimization or model distillation.

These findings validate CLIP as a powerful tool for creativity assessment, particularly in tasks requiring integration of visual and textual semantics. The observed improvements in originality and elaboration predictions highlight CLIP's strength in capturing abstract creative qualities. While text-only models retain an edge in title-related tasks, CLIP provides a unified framework suitable for general multimodal prediction. The successful integration into an interactive Streamlit application further demonstrates its practical utility for real-time creativity evaluation in both educational and clinical settings.

5 Conclusions

This study demonstrates the effectiveness of CLIP-based multimodal embeddings for automated creativity assessment across four key dimensions: originality, elaboration, flexibility, and title creativity. By integrating visual and textual representations, the proposed approach outperformed previous image- and text-only baselines, particularly in the evaluation of originality and elaboration. The Random Forest classifier leveraging raw CLIP embeddings achieved a new state-of-the-art AUC of 0.92 for originality prediction, while the MLP regressor slightly improved elaboration prediction.

Although the CLIP-based models did not surpass specialized language models like BETO in title creativity assessment, they maintained competitive performance, highlighting the complementary role of visual context in creative evaluation. Flexibility predictions using CLIP also yielded balanced metrics, indicating that fused embeddings are capable of capturing thematic diversity without relying solely on linguistic cues.

The successful deployment of these models in an interactive Streamlit web application further validates their practicality for real-time creativity evaluation in educational and clinical settings. Users can receive instant feedback on their creative work, fostering engagement and supporting personalized interventions.

Overall, this research advances the field of AI-assisted creativity assessment by introducing a scalable, interpretable, and user-facing framework that bridges cognitive science and machine learning. Future work may explore multilingual fine-tuning of CLIP, interpretability enhancements, and expansion to larger and more diverse datasets.

References

1. Acar, S., Organisciak, P., Dumas, D.: Automated scoring of figural tests of creativity with computer vision. The Journal of Creative Behavior **59**(1), e677 (2025). https://doi.org/https://doi.org/10.1002/jocb.677, https://onlinelibrary.wiley.com/doi/abs/10.1002/jocb.677
2. Allen, T.E., Chen, M., Goldsmith, J., Mattei, N., Popova, A., Regenwetter, M., Rossi, F., Zwilling, C.: Beyond Theory and Data in Preference Modeling: Bringing Humans into the Loop, p. 3–18. Springer International Publishing (2015). https://doi.org/10.1007/978-3-319-23114-3_1
3. Bahcecik, S.O.: I trends security politics and artificial intelligence: Key trends and debates. International Political Science Abstracts **73**(3), 329–338 (Jun 2023). https://doi.org/10.1177/00208345231182638, http://dx.doi.org/10.1177/00208345231182638
4. Beaty, R.E., Johnson, D.R.: Automating creativity assessment with semdis: An open platform for computing semantic distance. Behavior Research Methods **53**(2), 757–780 (Aug 2020). https://doi.org/10.3758/s13428-020-01453-w, http://dx.doi.org/10.3758/s13428-020-01453-w
5. Cañete, J., Chaperon, G., Fuentes, R., Ho, J.H., Kang, H., Pérez, J.: Spanish pre-trained bert model and evaluation data. In: PML4DC at ICLR 2020 (2020)
6. Cetinic, E., She, J.: Understanding and creating art with ai: Review and outlook (2021). https://doi.org/10.48550/ARXIV.2102.09109, https://arxiv.org/abs/2102.09109
7. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1800–1807 (2016), https://api.semanticscholar.org/CorpusID:2375110
8. Creely, E., Blannin, J.: The implications of generative ai for creative composition in higher education and initial teacher education. ASCILITE Publications p. 357–361 (Nov 2023). https://doi.org/10.14742/apubs.2023.618, http://dx.doi.org/10.14742/apubs.2023.618
9. Cropley, D.H., Marrone, R.L.: Automated scoring of figural creativity using a convolutional neural network. Psychology of Aesthetics, Creativity, and the Arts (Jul 2022). https://doi.org/10.1037/aca0000510, http://dx.doi.org/10.1037/aca0000510
10. Devedzic, V.: Is this artificial intelligence? Facta universitatis - series: Electronics and Energetics **33**(4), 499–529 (2020). https://doi.org/10.2298/fuee2004499d, http://dx.doi.org/10.2298/fuee2004499d

11. Ferrara, S., Qunbar, S.: Validity arguments for ai-based automated scores: Essay scoring as an illustration. Journal of Educational Measurement **59**(3), 288–313 (Jun 2022). https://doi.org/10.1111/jedm.12333, http://dx.doi.org/10.1111/jedm.12333
12. Franceschelli, G., Musolesi, M.: Deepcreativity: measuring creativity with deep learning techniques. Intelligenza Artificiale **16**, 151–163 (2022). https://doi.org/10.3233/ia-220136
13. Gigi, A.: Human figure drawing (hfd) test is affected by cognitive style. Clinical and Experimental Psychology **02** (2015). https://doi.org/10.4172/2471-2701.1000111
14. Harré, M.S., El-Tarifi, H.: Testing game theory of mind models for artificial intelligence. Games **15**(1), 1 (Dec 2023). https://doi.org/10.3390/g15010001, http://dx.doi.org/10.3390/g15010001
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015), https://arxiv.org/abs/1512.03385
16. Imuta, K., Scarf, D., Pharo, H., Hayne, H.: Drawing a close to the use of human figure drawings as a projective measure of intelligence. PLoS ONE **8** (2013). https://doi.org/10.1371/journal.pone.0058991
17. Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., Mikolov, T.: Fasttext.zip: Compressing text classification models. arXiv preprint arXiv:1612.03651 (2016)
18. Lee, S.W., Kwak, D.S., Jung, I.S., Kwak, J.H., Park, J.H., Hong, S.M., Lee, C.B., Park, Y.S., Kim, D.S., Choi, W.H., Ahn, Y.H.: Partial androgen insensitivity syndrome presenting with gynecomastia. Endocrinology and Metabolism **30**(2), 226 (2015). https://doi.org/10.3803/enm.2015.30.2.226, http://dx.doi.org/10.3803/enm.2015.30.2.226
19. Pikatza-Huerga, A., Matanzas de Luis, P., Uribe, M., Lasa, J., Zulaika, U., Almeida, A.: Analysing the impact of images and text for predicting human creativity through encoders. In: Proceedings of the 11th International Conference on Information and Communication Technologies for Ageing Well and e-Health. pp. 15–24. SCITEPRESS - Science and Technology Publications (2025)
20. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I.: Learning transferable visual models from natural language supervision (2021), https://arxiv.org/abs/2103.00020
21. Røed, R.K., Baugerud, G.A., Hassan, S.Z., Sabet, S.S., Salehi, P., Powell, M.B., Riegler, M.A., Halvorsen, P., Johnson, M.S.: Enhancing questioning skills through child avatar chatbot training with feedback. Frontiers in Psychology **14** (Jul 2023). https://doi.org/10.3389/fpsyg.2023.1198235, http://dx.doi.org/10.3389/fpsyg.2023.1198235
22. Searle, J.R.: Minds, Brains and Programs, p. 18–40. Routledge (May 2018). https://doi.org/10.4324/9781351141529-2, http://dx.doi.org/10.4324/9781351141529-2
23. Shaban-Nejad, A., Michalowski, M., Bianco, S., Brownstein, J.S., Buckeridge, D.L., Davis, R.L.: Applied artificial intelligence in healthcare: Listening to the winds of change in a post-covid-19 world. Experimental Biology and Medicine **247**(22), 1969–1971 (Nov 2022). https://doi.org/10.1177/15353702221140406, http://dx.doi.org/10.1177/15353702221140406
24. Sheng, L., Yang, G., Pan, Q., Xia, C., Zhao, L.: Synthetic house-tree-person drawing test: A new method for screening anxiety in cancer patients. Journal of Oncology **2019** (2019). https://doi.org/10.1155/2019/5062394

25. Sun, J., Zhang, J., Chen, Q., Yang, W., Wei, D., Qiu, J.: Psychological resilience-related functional connectomes predict creative personality. Psychophysiology **61** (2023). https://doi.org/10.1111/psyp.14463
26. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision (2015), https://arxiv.org/abs/1512.00567
27. Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks (2020), https://arxiv.org/abs/1905.11946
28. Weidinger, L., Reinecke, M.G., Haas, J.: Artificial moral cognition: Learning from developmental psychology (Aug 2022). https://doi.org/10.31234/osf.io/tnf4e, http://dx.doi.org/10.31234/osf.io/tnf4e
29. Zhang, R., Zeng, B., Yi, W., Fan, Z.: Artificial Intelligence Painting: A New Efficient Tool and Skill for Art Therapy (3 2024). https://doi.org/10.3233/FAIA240087