Journal of Inclusive Methodology and Technology in Learning and Teaching
ISSN 2785-5104
Anno3 n.4(2023)

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

EMULATION AND UNDERSTANDING THE EMOTION ACCORDING TO GENERATIVE ARTIFICIAL INTELLIGENCE - CASE STUDY OF EMOTIONAL COMPONENT EXTRACTED FROM VISUAL ARTWORKS

Umberto Bilotti[1], Lucia Campitiello[1], Michele Domenico Todino[1] and Maurizio Sibilio[1]

[1] University of Salerno; ubilotti@unisa.it; lcampitiello@unisa.it; mtodino@unisa.it; msibilio@unisa.it

Abstract: Artificial Intelligence can emerge as a new generative force in educational technologies, particularly through Generative Artificial Intelligence (GAI), making the production of various types of digital content faster and more accessible than ever. For the proper implementation of these technologies in the educational context, multidisciplinary and shared reflection is necessary. This article explores current GAI methods for image creation and proposes a study on the machine's current capabilities in interpreting and emulating human emotional phenomena. The emotional component extraction from visual artworks is chosen as a case study, and several possible interventions are identified.

Keywords: Generative Artificial Intelligence; text-to-image; AI learning.

1. Introduction

The aspect that has the greatest impact on society when a new technology is introduced is certainly its versatility. Society undergoes a change proportionated to the number and size of contexts where the application of that particular technology becomes easy and effective to simplificate human work. Based on these observations, we can identify Artificial Intelligence (AI) as one of the main factors capable of proposing an alternative to multiple processes that have traditionally distinguished between human and machine capabilities. Generative Artificial Intelligence (GAI) is a specific type of AI capable of generating several kinds of original content. As result from work of Jovanovic & Campbell (2022), there are currently several GAIs that specialize in producing specific types of content, such as text, image, audio, video programming code, etc. The production of this digital content has already reached a level of quality to satisfy certain demands of various fields including engineering, biology, medicine, and environmental sciences. For instance, GAI holds significant potential to impact the field of educational re-search for several compelling reasons. Firstly, its ability to autonomously generate content, such as educational materials and assessments, can streamline and enhance the creation process. This efficiency not only reduces the burden on educators but also facilitates the rapid development of

diverse and customized learning resources tailored to individual student needs this can take place for example in immersive educational virtual reality environments for distance learning workshops (Di Tore S. et al. 2022). However, in this paper we will look at general aspects of this technology that will later be included in the broader educational research of the 'Teaching Learning Centre for Education and Inclusive Technologies - Elisa Frauenfelder' lab (Labh.it/DISUFF). Therefore, when the objectives become more specific, the adoption of this technology necessitate a level of reliability that requires human control during the creation process. Unfortunately, to date, this interaction is not always feasible and easy or free for many of the users who have already experienced the potential offered by GAI.

Hence, in this paper we will explore one of the most interesting cases of artificial production such as image generation. In fact, in the first part of this contribution, we will focus on potential interventions for users who are tasked with assuming the role of an educator-teacher of AI. In the second part, we will employ the understanding and emulation of the emotional component in visual artworks, with a focus on GAI, as a case study. Moreover, we will also offer recommendations for dealing with critical issues that have emerged from the study.

2. Images from Generative Artificial Intelligences

Unlike some forms of synthetic data, the production of which is intuitive and accessible to everyone, such as text and, in some cases, images, others remain the domain of experts or require more powerful machines on average than those commonly found, especially in the various laboratories or classrooms of primary and secondary schools. For instance GauGAN, has the capability to transform very simple sketches into synthetic panorama images in real-time (Park et al. 2019). Its usage through NVIDIA Canvas software may be easily understood even by very young age groups, with surprising results. However, requirements for installation demand very powerful GPUs. Its extension, GauGAN2, also enables the creation of panorama images through keyboard-typed prompts. This allows for the initial image to be generated based on a short description with additional details added by drawing them similar to its previous version. This mode of image creation, which takes as input text typed in by the user, is certainly attentioned by several researchers. It has led to the development of popular tools like the three successive versions of DALL-E developed by OpenAI for the creation of synthetic images and artistic images (A. Ramesh et al. 2021, 2022). Another text-to-image method proposed by GoogleAI is Imagen (Saharia C. et al 2022); which is specialized in the creation of realistic high-resolution images. Of particular interest for educational purposes is Imagen Editor, which allows separate regions of the same image to be edited, through multiple prompts (Wang S. et al 2023). In contrast to the other image generators that offer web apps or downloadable software, using Midjourney requires the subscription to the Discord channel, joining one of the image-generating servers, and sending an '/image' message in the chat, you to initiate the prompt for entering the description of the target image. The text image generator provided by Stable Diffusion 1.5 is undoubtedly one of the simplest to use, as users only needs to directly input the prompt into the web app and select the image style from the drop-down menu. Although the picture style selection may seem somewhat limiting, the results obtained are highly convincing. In fact, the average accuracy in distinguishing synthetic images from genuine ones is only 46.97%, as demonstrated by Zhang J. et al. (2023).

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

The growing interest in these technologies, as along with the availability of aids such as tutorials or manuals for effective usage, has already led to the development of specific courses like the 'AI and Art' course offered at Quinnipiac University, as described in the work of G. Patrick (2023). The shared goal of producing high-quality images and derivatives has intensified interest in the engineering of structures made of several GAIs. The framework proposed in the work of Guo C. et al. (2023) for creating digital images not only contemplates the integrations of two AIs (one for generating a more efficient prompt and the other for creating the image), but also effectively identifies how human intervention may contribute to progressive and iterative improvement of the result. In fact, human intervention is valuable neither only in the technical design of the AI algorithm, nor only at the detailed and coherent structuring of the data set for the training phase of the algorithm, but also in providing feedback on the output generated by the machine.

Figure 1. An example of a landscape created with NVIDIA Canvas.

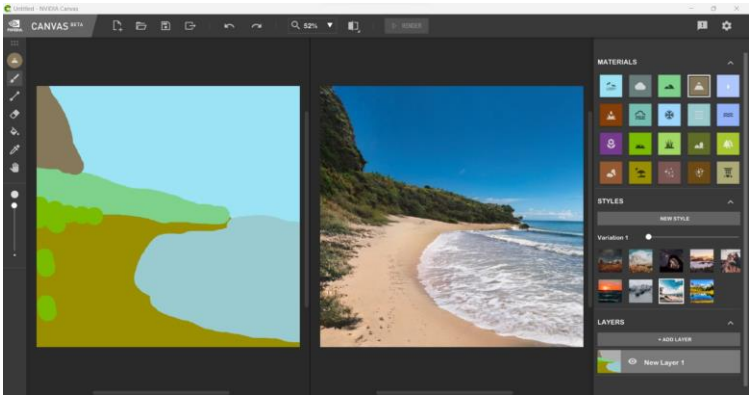


Figure 2. In the image the different results from a strongly synthetic prompt as "A view of the Salerno coast, sea, night, Arechi Castle, Van Gogh." and a reformulation of the prompt made by ChatGPT 3.5 as "Imagine a night on the coast of Salerno, with the sea reflecting the moonlight. The imposing Arechi Castle on the hill stands against the starry sky. City lights dance on the waves while the sea breeze carries the salty scent in the air. In this scene, you could envision elements inspired by the style of Van Gogh, with distinct painterly strokes capturing the nocturnal atmosphere and the seascape in a vibrant palette of colors"

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

3. Text-to-image as a new way of learning by teaching AI and vice versa.

The framework described in the final part of the previous paragraph is crucial for hypothesizing the extension of a common pedagogical-didactic approach to intelligent machines. The underlying assumption is that the two processes of understanding and generating information are mutually symbiotic. In fact, the collaboration between multiple GAIs and human intelligence not only results in a higher-quality product, but also enhances the machine's comprehension of the product. The prospect that communication between humans and machine can take place not in a unidirectional way and not limited to ICT experts alone may bring about the scenario where interdisciplinary fields contaminate each other in a synchronous and reciprocal way. This allows for a space for dialogue and project development through a collaborative negotiation of models, intents, actions, practices, as well as effective and ethical outcomes, in order to promote more participatory and concrete approaches(Panciroli C. et al. 2020).In more detail, an AI system can integrate the Diversity, Equity, Access, and Inclusion, approach (DEAI; Todino, M. D. 2023) by adopting a multifaceted strategy that covers several dimensions. Implementing DEAI principles involves designing and training the AI model with diverse datasets that accurately represent different demographic groups. This ensures that the AI system is not biased towards any particular group and provides a fair and inclusive representation.

Hence, beyond the creation of software like Machine Learning for Kids, eCraft2Learn, Cognimates, etc. (Hsu T. et al. 2021), designed to enhance the comprehension of AI concepts among primary and secondary school students, the efficacy of AI tools in educational settings is reinforced by a robust adaptive component. This adaptability is crucial in facilitating the customization of the learning experience according to the specific requirements of end-users, making the adaptation process straightforward and user-friendly. Communication between humans and machines, already widely 'naturalised' thanks to established fields of research such as Human-Computer-Interaction as well as current Natural Language Processing methods such as ChatGPT, may gain a new mode of expression with text-to-image methods. The possibilities offered by these new technologies may be explored in the established but evolving pedagogical framework. As described in the work of Di Tore S. (2016), in a society permeated by technology, schools should not resist it, but instead embrace it, utilizing it to enhance teaching practices rather than replacing them. According to this principle, the teacher's role is pivotal in channeling the functionality of technology as a deviant strategy for more inclusive and ecologically cohesive learning. The introduction of these text-to-image applications aligns with this perspective. In this particular case, we should consider whether traditional artistic and iconographic reading techniques are still effective in order to prevent AI from hindering students' creative processes. However, as Luigini A. (2023) points out, since the machine's image creation process follows mathematical rules, the student's poietic process is unaffected because creativity is observed in the choice of text prompts, manipulation of the images using other specific software, and continuous validation of the subsequent resulting images. The production of visual content through AI retains its educational potential when it stems from an alternative instrumental knowledge system that may be structured and oriented to a specific purpose. For example, very tangible tools like brushes and pencils can find valid alternatives in computer mouses or joypads of immersive viewers when the study of painting style is combined with a study of image

editing software, just as image composition may be efficiently achieved through AI when its components is combined with an appropriate prompt.

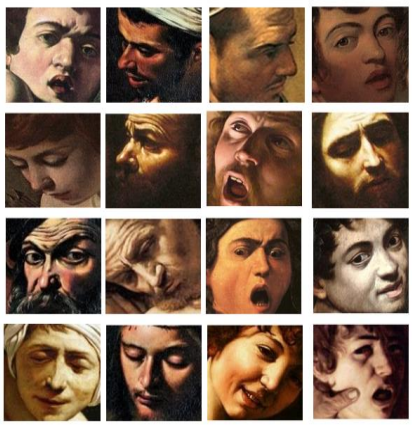
4. AI understanding and analysis of the emotional component in visual artworks and possible interventions

Once we overcome initial concerns about the effectiveness of using these technologies, unfortunately, the observations made so far alone may not be sufficient when we add the constraint that a certain visual production should trigger or encapsulate an emotional expression. Visual art, deeply woven into the fabric of human history, emerges as a form of expression of profound relevance and representation for various cultures worldwide. These visual creations provide a unique insight into how different societies interpret and communicate the complexity of their world through visual language. The accurate interpretation of the expressive potential inherent in each of these artworks becomes a valuable reading key, a means access profound layers of human experiences. If well-executed, their artificial reworking not only preserves the aesthetic beauty but may also serves as a kind of emotional laboratory. Through this process, artworks become mirrors of human emotions, providing a framework to explore and understand the subtle nuances of the human psyche.
However, despite the abundance of data derived from centuries of artistic tradition, the currently developed methods to detect or emulate emotional stimuli through these artworks may not fully harness their potential. One potential difficulty involves the datasets used to train Generative Artificial Intelligences (GAIs), which may not yet have enough high-quality material. Training on AI models of datasets densely packed with information relevant to a specific goal should be part of creating an AI suited for a particular task. Initial observations of available datasets of visual artworks indicate that some were created for different purposes, despite their large size. Therefore, they may also include artworks devoid of human faces or emotionally challenging to interpret (examples include landscapes, still life, abstract art, and so on). Other datasets more aligned with our goal, on the other hand, have fewer samples or a less representative classification of emotional phenomena. Humans have historically communicated the majority of their emotional state through body language, with a strong emphasis on facial expressions. To reliably identify a specific facial expression linked with a particular emotional spectrum, the face must first be extracted from an image. Most modern Face Detection algorithms that use AI have been trained on images, typically taken under controlled lighting and position conditions, which are uncommon in the case of visual artworks. Several research have also shown that the display of an emotional experience is not limited to facial expressions or body language. Simpler elements, such as geometric patterns, abstract forms, lines, and colors, can both evoke and result from an emotional state. In other words, emotion may be seen in a wider spectrum of visual features than only human expressions, paving the way for a more comprehensive approach to emotional analysis of visual artworks.

Table 1. Some datasets containing images visual artwork.

Dateset	Work	Number of images	Type of data	Use
Art500k	Mao, H. et al. 2017	554198	Low resolution images of visual artworks	Contents and style analysis of visual arts
ArtEmis	Achlioptas, P. et al. 2021	80031	Portraits, landscapes, abstract art	Painting captioning
WikiArt Emotions	Mohammad, S., et al. 2018	4105	Paintings with/without the presence of a face and with/without the presence of a body	Future works
MART	Yanulevskaya, V. et al. 2012	500	Compositions of lines, colours, shapes and tex-tures	Emotion recognition
No name	Luo, S et al. 2022	114	Portraits and natural landscapes	Stimulus to detect the viewer's emotion
No name	Lu, Y et al. 2023	9869	Portraits and paintings with human subjects	Painting captioning

Figure 3. For the extraction of these faces, the Face Detection method had to be improved by adding rotations of 90°, 180° and 270° for each input image.



5. Conclusions

Expanding the horizons of academic research in the field of GAI and education, as mentioned before, one promising avenue lies in advancing the integration of GAI to enhance personalized learning. Harnessing the power of generative algorithms, researchers can explore the development of adaptive learning systems that customize educational content to individual students based on their unique learning styles and preferences. This approach has the potential to revolutionize the education sector by offering a more personalized and engaging learning experience, ultimately improving student outcomes. By leveraging its capabilities in data analysis, GAI can expedite the research workflow, allowing educators and researchers to concentrate on higher-level tasks such as result interpretation and the formulation of innovative pedagogical methodologies. This could lead to more efficient and effective educational research practices, driving advancements in teaching-learning process. Moreover, to align with the DEAI principles, future research endeavors could explore ways to embed diversity, equity, access, and inclusion considerations within GAI systems. This involves addressing potential biases in the algorithms used for personalized learning and ensuring that the generated content is inclusive and accessible to diverse student populations. Research in this area can contribute to the development of AI technologies that not only enhance education but also adhere to ethical and inclusive standards.

References

Achlioptas, P., Ovsjanikov, M., Haydarov, K., Elhoseiny, M., & Guibas, L. J. (2021). Artemis: Affective language for visual art. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11569-11579).

Di Tore, S. (2016). La tecnologia della parola. Didattica inclusiva e lettura. FrancoAngeli.

Di Tore, S., Campitiello, L., Caldarelli, A., Todino, M. P., Di Tore, P. A., Iannaccone, A. & Sibilio, M. (2022). EDUCATION IN THE METAVERSE: AMIDST THE VIRTUAL AND REALITY L'EDUCAZIONE NEL METAVERSO: TRA VIRTUALE E REALE. https://www.researchgate.net/publication/367092002

Guo, C., Lu, Y., Dou, Y., & Wang, F. Y. (2023). Can ChatGPT boost artistic creation: The need of imaginative intelligence for parallel art. IEEE/CAA Journal of Automatica Sinica, 10(4), 835-838.

Hsu, T. C., Abelson, H., Lao, N., Tseng, Y. H., & Lin, Y. T. (2021). Behavioral-pattern exploration and development of an instructional tool for young children to learn AI. Computers and Education: Artificial Intelligence, 2, 100012.

Jovanovic, M., & Campbell, M. (2022). Generative artificial intelligence: Trends and prospects. *Computer*, *55*(10), 107-112.

Lu, Y., Guo, C., Dai, X., & Wang, F. Y. (2023). Generating Emotion Descriptions for Fine Art Paintings via Multiple Painting Representations. *IEEE Intelligent Systems*.

Luigini, A. (2023). AI imaging, imagery and imagination: Considerations on a future that is already present, for a digital humanism in poietic and educational processes. In IMG23: Atti del IV Convegno Internazionale e Interdisciplinare su Immagini e Immaginazione= Proceedings of 4th International and Interdisciplinary Conference on Images and Imagination (pp. 336-343). Publica.

Luo, S., Lan, Y. T., Peng, D., Li, Z., Zheng, W. L., & Lu, B. L. (2022, July). Multimodal emotion recognition in response to oil paintings. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 4167-4170). IEEE.

Mao, H., Cheung, M., & She, J. (2017, October). Deepart: Learning joint representations of visual arts. In Proceedings of the 25th ACM international conference on Multimedia (pp. 1183-1191).

Mohammad, S., & Kiritchenko, S. (2018, May). Wikiart emotions: An annotated dataset of emotions evoked by art. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*.

Panciroli, C., Rivoltella, P. C., Gabbrielli, M., & Richter, O. Z. (2020). Artificial Intelligence and education: new research perspectives Intelligenza artificiale e educazione: nuove prospettive di ricerca. Form@ re-Open Journal per la formazione in rete, 20(3), 1-12.

Park, T., Liu, M. Y., Wang, T. C., & Zhu, J. Y. (2019). Gaugan: semantic image synthesis with spatially adaptive normalization. In ACM SIGGRAPH 2019 Real-Time Live! (pp. 1-1).

Patrick, G. (2023, July). Artificial Intelligence and Art: A University Curriculum Course for Undergraduates. In 2023 14th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI) (pp. 667-670). IEEE.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. arXiv preprint arXiv:2204.06125, 1(2), 3.

Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., ... & Sutskever, I. (2021, July). Zero-shot text-to-image generation. In International Conference on Machine Learning (pp. 8821-8831). PMLR.

Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E. L., ... & Norouzi, M. (2022). Photorealistic text-to-image diffusion models with deep language understanding. Advances in Neural Information Processing Systems, 35, 36479-36494.

Todino M.D., (2023) Accessibility in museums: how to combine accessible spaces, technology, and "shared languages" to ensure the inclusion of all visitors, Giornale Italiano di Educazione alla Salute, Sport e Didattica Inclusiva - Italian Journal of Health Education, Sports and Inclusive Didactics. Anno 7, V 3. Edizioni Universitarie Romane

Wang, S., Saharia, C., Montgomery, C., Pont-Tuset, J., Noy, S., Pellegrini, S., ... & Chan, W. (2023). Imagen editor and editbench: Advancing and evaluating text-guided image inpainting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 18359-18369).

Yanulevskaya, V., Uijlings, J., Bruni, E., Sartori, A., Zamboni, E., Bacci, F., ... & Sebe, N. (2012, October). In the eye of the beholder: employing statistical analysis and eye tracking for analyzing abstract paintings. In Proceedings of the 20th ACM international conference on multimedia (pp. 349-358).

Zhang, J., Wang, Y., Tohidypour, H. R., & Nasiopoulos, P. (2023, October). Detecting Stable Diffusion Generated Images Using Frequency Artifacts: A Case Study on Disney-Style Art. In 2023 IEEE International Conference on Image Processing (ICIP) (pp. 1845-1849). IEEE.