Iraqi Journal for Computer Science and Mathematics

Journal Homepage: http://journal.esj.edu.iq/index.php/IJCM
e-ISSN: 2788-7421 p-ISSN: 2958-0544

Development of synthetic data generator for ornament based on data mining techniques

Saad Ahmed Dheyab[1], Rasha Talal Hameed[2], Ahmed Hussein Ali[1]*

[1]College of Engineering, University of Information Technology and Communications, Iraq.
[2]College of Education, Al-Iraqia University, Iraq.

*Corresponding Author: Ahmed Hussein Ali

DOI: https://doi.org/10.52866/ijcsm.2024.05.03.045
Received April 2024 ; Accepted June 2024; Available online September 2024

ABSTRACT: Artificial intelligence tools depend on generating various images on the available datasets, but there is a lot of data that is not completely available, especially images of heritage and archaeological ornament. In this paper, a tool was developed to generate ornament based on the geometric shapes of heritage ornaments. A group of basic geometric shapes was collected. The new ornament are reshaped in three stages, relying on the permutation equation and reshaping of the overlapping sizes of the shapes, In the first stage, the number of layers that make up the ornament is determined. In the second stage, the formation of the ornament is determined through rotation or mirroring. In the final stage, the set of ornaments is generated in the form of an integrated matrix.

Keywords: ornament, synthetic data, data mining, artificial intelligence tool.

1. INTRODUCTION

Ornamentation consists of a group of interlocking points, lines, and shapes that give distinctive geometric shapes used to decorate buildings. Ornamentation is one of the arts sciences that aims to research the philosophy of proportion, abstraction, mass, space, composition, line, and color [1]. The Ornamentation are of various types, including botanical ornament, which relies on the leaves of plants and trees in its formation, written ornament that rely on letters to form the decoration, which is often used in the Arabic language, geometric ornament that rely on geometric shapes such as squares, triangles, circles, and lines in formation, and pictorial ornament that relies on objects Living in formation[2]. The large number of possibilities for overlapping geometric shapes and forming decoration has become a necessity in using smart tools to generate all these possibilities of ornament that help artists or those interested in forming and studying ornament [3].

2. Image mining techniques

Image mining is based on extracting tacit knowledge and patterns not explicitly stored in images through image thresholding, edge detection and boundary tracking followed by association rule mining [4]. The techniques used in image mining are the same as those used in data mining including image classification, image clustering, and association rules in images. Image mining includes a set of levels of operations as follows [5]:

a- Level of knowledge extraction.
b- Level of patterns relations.
c- Level of semantic concept.
d- Level of visual patterns

3. AI image generators

Many institutions and researchers have sought to develop artificial intelligence tools to generate images for commercial and research purposes. These tools have varied in their generation mechanism. Some rely on descriptive texts to generate the image, while others rely on images to generate images. The most prominent challenge faced by these tools is the accuracy of the details of the resulting image and the extent of their matching for the descriptive data, among the most famous tools that rely on natural processing language and text processing based on deep learning are Diffusion, DALLE, Craiyon, Stable, Glide and Midjourney [6,7].

4. Permutations and Combinations

It is the process of arranging the elements of a group in a series or a specific order. If the elements are arranged, the process of rearranging their elements is called permutation. Permutations differ from combinations, which are defined as selections of elements from a set without regard to order, it is defined by the formula [8]:

$$nPr = \frac{n!}{(n-r)!} \qquad (1)$$

Where repetition is not allowed, r objects are chosen from n different distinct objects. If n = 3 and r = 3, according to the formula, we will have 6 groups of permutations. Note Fig.1.

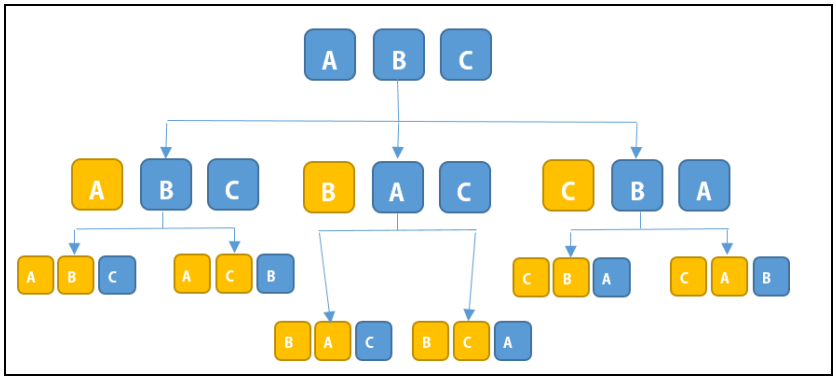

FIGURE 1. Example of permutations

5. Synthetic Data

Synthetic data is defined as the data that is generated through computer algorithms, this method of generation is relied upon to obtain data that is not sufficiently available in real world to enhance artificial intelligence algorithms and train them more on this data and enable them to generate more effective data in the health sectors, commercial and industrial[9], One of the advantages of using this type of data is that it breaks away from the limitations imposed by the available data, which makes it more effective in software testing and ensuring quality, The challenge with this data remains the inconsistencies that may occur and the inability to generate accurate data, Fig. 2 shows the generation of synthetic data [10].

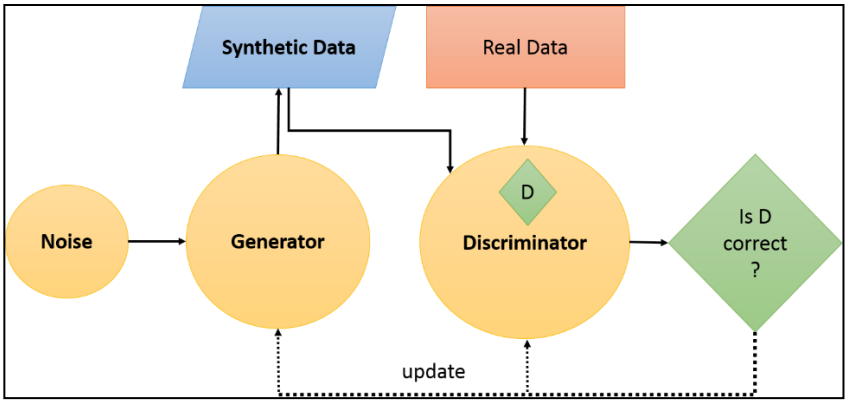

FIGURE 2. Generation of synthetic data

6. Characteristics of synthetic data

The quality of data, patterns, and trends are the most important thing that concerns data scientists, as they do not consider whether the data is real or synthetic. Therefore, synthetic data must have some characteristics so that data scientists can benefit from it, and these characteristics include [11]:

1. Data quality: The high cost and sources of obtaining data from the real world may lead to some errors, biases and inaccuracies in data, so institutions that rely on synthetic data to train artificial intelligence models seek to focus on improving the quality of data.

2. Scalability: The increasing demand for training data has made real data insufficient, so using synthetic data and making it scalable is the ideal solution to meet the training needs of artificial intelligence models [12].
3. Simple and effective: Synthetic data must be simple in its generation mechanism, but it must be effective and not contain any biases or inaccuracies that could negatively affect training [13].

7. Types of synthetic data

Due to the sensitivity of using Synthetic data in training, it is necessary to know the type that suits the training model. Synthetic data is divided into two parts:

1. Fully synthetic data: All required variables are available in this type, but data has no connection to real data. This type of data is used when there is no real data, but there are specifications and limitations for data [14].
2. Partially synthetic data: Data in this type retains all real data except sensitive data or data containing errors or biases that are deleted and replaced with the synthetic dataset [15].

8. Synthetic data generation methods

Synthetic data relies on many methods to generate it, taking into account the characteristics and limitations, as follows:

1. Statistical method: statistical distributions are relied upon to extract numbers and values from real data and reproduce realistic data through the normal distribution, exponential distribution, and other distributions [16].
2. Agent to model method: In this method, an agent based on machine learning algorithms is relied upon to generate synthetic data that is very similar to real data and is more effective than traditional statistical methods [17].
3. Deep learning: In this method, some deep learning models are used that ensure that the inputs are very similar to the outputs and from unsupervised machine learning such as variational autoencoder or generative adversarial model [18].

9. Proposed work

A three-stage approach was proposed to build a smart tool to generate synthetic data represented by a group of heritage ornament. The initial processing stage consists of collecting a group of ornaments, extracting the basic geometric shapes from each ornament, and isolating them as independent images so that new ornaments bearing the same pattern can be reshaped, the basic shapes were addressed through the three stages, as follows:

1. First stage: This stage works on choosing the basic shapes and combining them in the form of layers of different sizes and using the permutation equation to form all the overlapping possibilities for the layers and form the basic unit of the ornament, as shown in Fig. 3.

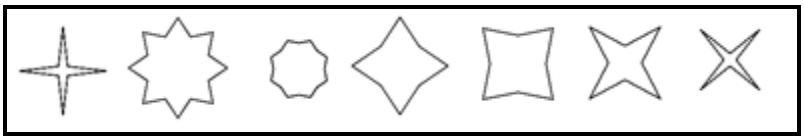

FIGURE 3. Basic shapes.

In Figure 3, 7 basic shapes have been identified, which represent value of N in the permutation equation. Here, the value of R must be determined, which represents the number of overlapping layers to form the basic unit of ornament, as shown in Fig. 4.

Table 1 shows the number of basic ornaments generated according to the number of layers that are selected in each generation process, Increasing the number of layers generates more ornaments, which enhances the scalability of the resulting data within the parameters of the basic shapes. Generating the basic ornaments depends on a set of steps that can be summarized as follows:

Algorithm 1: basic ornaments generation.
Begin
Step1: Extracting basic shapes from real ornaments.

```
Step2: convert basic shapes into PNG files.
Step3: select no. of basic shapes as K. and no. of layers as r
Step4: make all PNG files as a same height and width.
Step5: loop from 1 to k
        Loop from 1 to r
        Call permutation method (k, r) (to select PNG files)
        Call resize image layer method as l
        Put PNG image in the layer l
        Resize layer * 1.5 (to Increase the scale of the next layer)
       Save the result
       End loop
    End loop
End
```

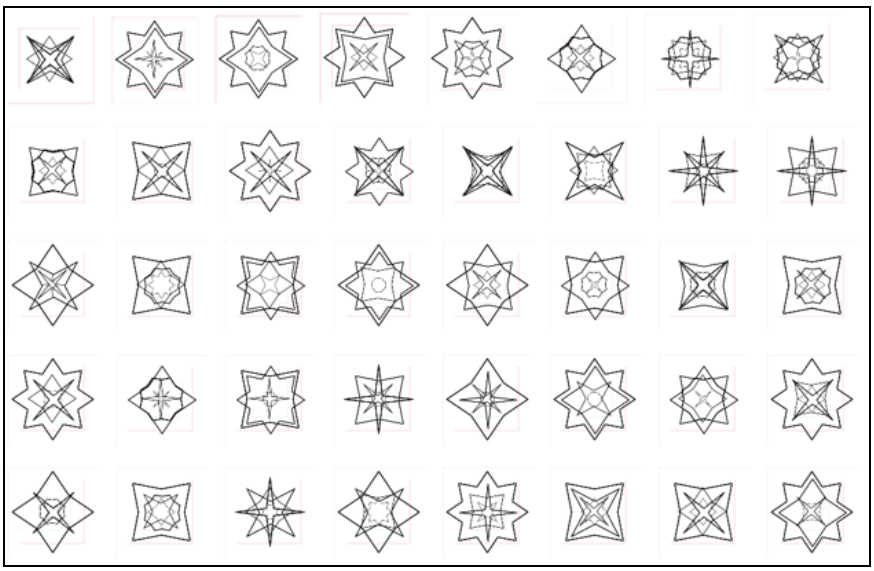

FIGURE 4. Samples of basic generated ornaments

2.Second stage: In this stage, the compositional ornament is formed from the group of ornaments that were generated in the first stage. This formation process takes place through the circular pattern or the reflective pattern, as shown in the fig. 5, circular pattern depends on placing a basic ornament in the middle of image and it is surrounded in a circular manner by an ornament of a different type. All possibilities are generated to link any ornament generated in the first stage with all other ornaments, which allows generation of various forms of data while preserving the basic pattern of ornament, The reflection pattern is done by constructing a specific pattern using two types of ornaments and then performing the reflection process. This is a very famous pattern in building ornaments, as shown in the fig. 6.

Table 1. Number of basic ornaments

Basic Shapes no.	No. of layers	No. of ornaments
7	2	42
7	3	210
7	4	840

8	2	65
8	3	336
8	4	1680

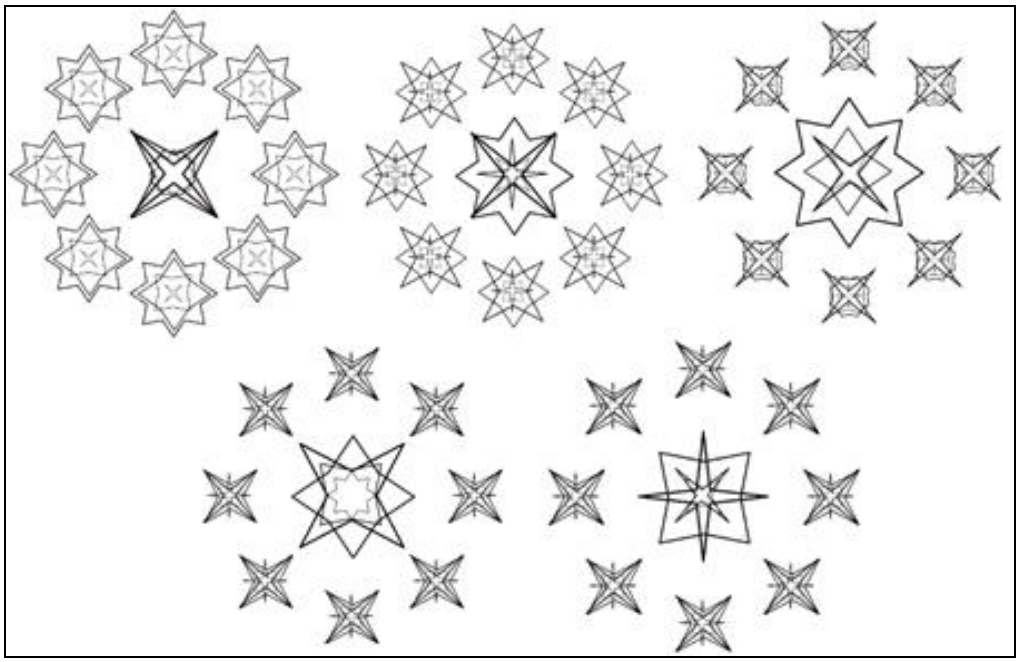

FIGURE 5. Samples of circular pattern ornaments.

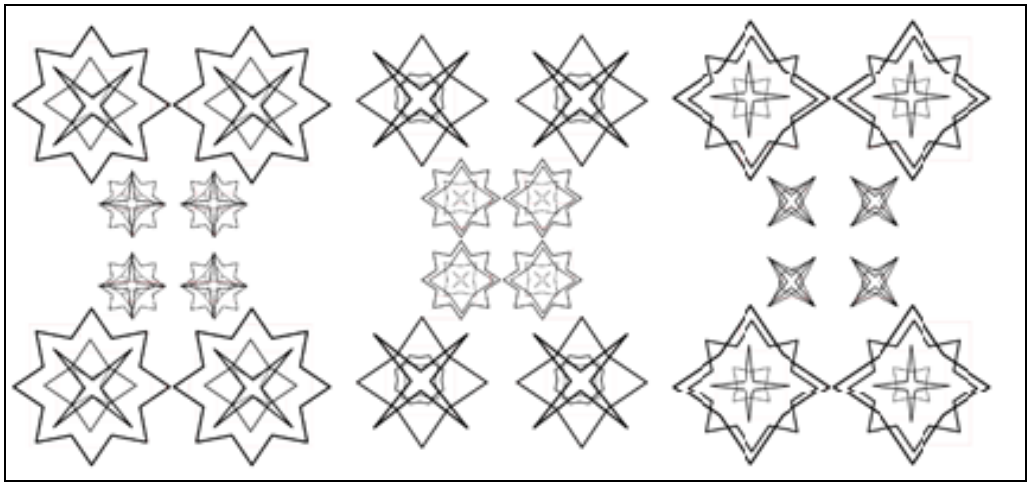

FIGURE 6. Samples of reflection pattern ornaments.

3.Third stage: The final stage is the stage of producing an integrated ornaments pattern by merging the patterns generated from the second stage, as shown in the figures 7 and 8.

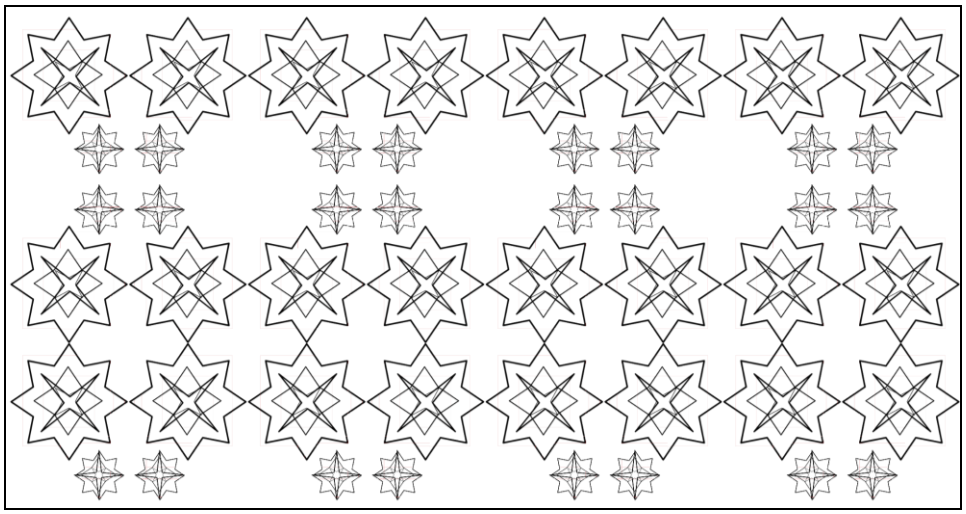

FIGURE 7. Sample of reflection integrated ornaments

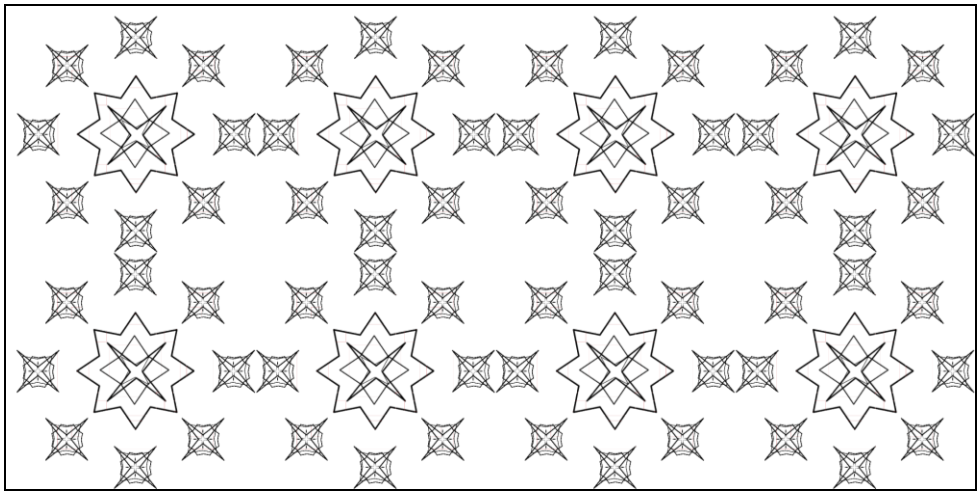

FIGURE 8. Sample of circular integrated ornaments

10. Conclusion and future work

This paper presents a proposed method for generating synthetic data for heritage ornaments based on the basic shapes that make up the original ornaments. Artificial intelligence requires more data for training, which consequently improves the ability of smart algorithms to recognize shapes, distinguish patterns, and reshape ornaments with specific patterns based on artificial intelligence. Generatively, the most important thing that can be done to develop the tool's work is to color the resulting ornaments and generate three-dimensional ornaments and shapes.

Funding

None

ACKNOWLEDGEMENT

None

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] Talia Ringer, Nathaniel Yazdani, John Leo, and Dan Grossman. Ornaments for Proof Reuse in Coq. In 10th International Conference on Interactive Theorem Proving (ITP 2019). Leibniz International Proceedings in Informatics (LIPIcs), Volume 141, pp. 26:1-26:19, Schloss Dagstuhl – Leibniz-Zentrum für Informatik (2019).

[2] Baker, Jack & Solange, Rigaud & Pereira, Daniel & Courtenay, Lloyd & d'Errico, Francesco. (2024). Evidence from personal ornaments suggest nine distinct cultural groups between 34,000 and 24,000 years ago in Europe. Nature Human Behaviour. 8. 10.1038/s41562-023-01803-6.

[3] Dabney, Mary & Wright, James. (2020). TOOLS, WEAPONS, FIGURINES, AND PERSONAL ORNAMENTS. 10.2972/j.ctv22wtqv8.19.

[4] Pandey, Sandeep & Sri Khetwat, Saritha. (2017). A Survey Paper on Image Classification and Methods of Image Mining. International Journal of Computer Applications. 169. 10-12. 10.5120/ijca2017914765.

[5] Chakraborty, Sayan & Abdessalem, Wahiba & Dey, Nilanjan & Banerjee, Sukanya & Salem, Mohammed Abdel-Megeed Mohammed & Azar, Ahmad. (2015). Image Mining Framework and Techniques: A Review. International Journal of Image Mining. In-press. 10.1504/IJIM.2015.070028.

[6] S. Göring, R. R. Ramachandra Rao, R. Merten and A. Raake, "Analysis of Appeal for Realistic AI-Generated Photos," in IEEE Access, vol. 11, pp. 38999-39012, 2023, doi: 10.1109/ACCESS.2023.3267968.

[7] Yadav, Nimesh Bali & Sinha, Aryan & Jain, Mohit & Agrawal, Aman & Francis, Sofia. (2024). Generation of Images from Text Using AI. International Journal of Engineering and Manufacturing. 14. 24-37. 10.5815/ijem.2024.01.03.

[8] Anekwe, Uchenna. (2019). Equations in Permutation and Combinations with Introduction of the Anekwe's Method of Swapping Factorials. Journal of Engineering and Applied Sciences. 4. 69-76.

[9] M. Xie, T. Hu and J. Zhou, Transfer learning under the Cox model with interval-censored data, Stat. Anal. Data Min.: ASA Data Sci. J. 17 (2024), e11680. https://doi.org/10.1002/sam.11680

[10] Fonseca, J., Bacao, F. Tabular and latent space synthetic data generation: a literature review. J Big Data 10, 115 (2023). https://doi.org/10.1186/s40537-023-00792-7

[11] Marcel Altendeitering, Tobias Moritz Guggenberger, Frederik Möller,A design theory for data quality tools in data ecosystems: Findings from three industry cases, Data & Knowledge Engineering,Volume153,2024,102333,ISSN0169023X,https://doi.org/10.1016/j.datak.2024.102333.

[12] Zhixin Li, Jiahui Wei, Feicheng Huang, Huifang Ma,Modeling graph-structured contexts for image captioning, Image and Vision Computing, Volume 129,2023,104591,ISSN 0262-8856,https://doi.org/10.1016/j.imavis.2022.104591.

[13] Flavio De Paoli, Michele Ciavotta, Roberto Avogadro, Emil Hristov, Milena Borukova, Dessislava Petrova-Antonova, Iva Krasteva, An interactive approach to semantic enrichment with geospatial data,Data & Knowledge Engineering,Volume 153,2024,102341, ISSN 0169-023X,https://doi.org/10.1016/j.datak.2024.102341.

[14] Saillant, B., Zara, F., Damiand, G. et al. High-order elements in position-based dynamics. Vis Comput 40, 4737–4749 (2024). https://doi.org/10.1007/s00371-024-03467-3

[15] Zhixin Li, Qiang Su, Tianyu Chen,External knowledge-assisted Transformer for image captioning,Image and Vision Computing,Volume 140, 2023,104864,ISSN 0262-8856,https://doi.org/10.1016/j.imavis.2023.104864.

[16] Towa Yamabe, Takeshi Saitoh, Chun-Ping Liu and Liang Xiao,Conference: Fifteenth International Conference on Graphics and Image Processing (ICGIP 2023), Year: 2024, Page 48,DOI: 10.1117/12.3021119.

[17] Bath, U., Shekhar, S., Egbert, J. et al. CERVI: collaborative editing of raster and vector images. Vis Comput 38, 4057–4070 (2022). https://doi.org/10.1007/s00371-022-02522-1

[18] Pise, Dr & Yadgiri, Naveen & Gaikwad, Preksha & Dusawar, Yashika & Nandanwar, Prathamesh. (2024). AI Image Generator. International Journal of Advanced Research in Science, Communication and Technology. 768-773. 10.48175/IJARSCT-18385.