Natural Language Driven Combinatorial Visual-Textual Presentation Layout Generation based on User-Centered Design Approach

by

Xiaolin Deng (she/her)

21001319

This thesis is submitted to University of the Arts London
in partial fulfilmentof the requirement for the degree of

MSc Data Science and AI for the Creative Industries

of

Creative Computing Institute

Supervisor: Dr Peaks Krafft

29 September 2022

The GitHub repo: https://git.arts.ac.uk/21001319/Combinatorial_Visual-Textual_Presentation_Layout_Generation

ual: creative computing institute

ABSTACT

This project addresses the possibility of taking a user-centered design (UCD) (Norman and Draper, 1986) approach to generate visual-textual presentation layouts based on natural language for designers and non-designers. By integrating semantic understanding, multi-modal learning, deep learning algorithms and graphic design principles, this project aims to support a combinatorial method of intelligence graphic design generation. Further, this project uses Chinese e-commerce graphic advertising, as exemplified, to focus on the generation of e-commerce advertising descriptions and intelligent graphic design. Using different machine learning models, the combinatorial generation approach provides an implementation path for the experiments. Although at this stage the results of the presentation layout are not ideal, this design theory and implementation path provide a new design thinking and discussion basis for designers and non-designers during the creative process, expanding the boundaries of their perception and imagination and better unleashing human creativity.

Keywords: UCD, multi-modal, deep learning, intelligent graphic design.

1 INTRODUCTION

Combined images with overlaid texts and embellishments, also known as visual-textual presentations (Yang et al., 2016), are becoming increasingly ubiquitous, as in the form of advertising posters and magazine covers. In this project, these presentations are referred to more generally as graphic design.

Most graphic designers spend little time 'creating' in the design process. In addition, their concept input and first output do not match well, requiring heavy adjustment to gradually improve the expected match level.At the same time, during the graphic design process, professional and non-professional designers face the problem of sorting out the demand of design and looking for inspiration before they can conceive a concept or produce a first design sketch. This also requires an understanding of the audience and the final application scenario of the graphic design you are working on.

Over the last few decades, several major technological breakthroughs have been made in deep learning research in areas such as computer vision (CV) and natural language processing (NLP). These techniques have made multimodal tasks one of the hot topics of research in the artificial intelligence (AI) industry, with the aim of building models that can process and correlate information from multiple modalities. Multimodal tasks are an increasingly important and dynamic multidisciplinary field with extraordinary potential. The creation of multimodal neural networks (Goh et al., 2021) is also considered a long-term goal of AI. This has led more researchers to focus on the textual and visual domains and has driven a change in traditional design methods. However, there is still a large gap between the research and practical applications of multimodal neural networks in the growing field of intelligent graphic design.

2 RELATED WORK

2.1 Trends in the integration of artificial intelligence and visual design

In recent years, AI has begun to be widely used in the field of visual creativity. The main roles played by AI in the visual design process can be divided into two categories, based on the support of the design process (Assistive Design) and the production of creative content (Automatic Design Generation) (Zhou et al., 2020). The former category is where the designer remains the primary provider of the design ideas, with AI acting as a design-assisting tool that facilitates end-user (Eric von Hippel, 2001) innovation in design and improves their complex creative processes. The latter refers to AI as the primary source of design ideas, focusing on generating creativity and collaborating with or even replacing the designer in the main design process. Innovative complementary AI tools have great potential to support design creation, as they can be applied to various design-related fields and broaden the design thinking process.At present, however, it is an emerging area of research with much room for development. There is a growing body of work examining the co-creative relationship between humans and AI in different fields.

2.2 Current status and trends in artificial intelligence-assisted visual design research

The specific empowerment of AI to participate in the traditional design process can be divided into four stages, namely Requirements Analysis, Inspiration Stimulation, Design Conceptualisation and Design Evaluation.

2.2.1 Requirements Analysis

User requirements analysis uses appropriate technical methods to understand as much as possible about the user and gain insight into their needs and wants. The success of design and creative work often depends on the designer's understanding and control of the user's needs. For example, AI processes large amounts of user data to achieve Automatic Persona Generation (APG) (Jung et al., 2018), IRGAN(Wang et al., 2017)and RecGAN (Bharadhwaj, Park and Lim, 2018) for generating relevant potential user preferences based on data samples of user behaviour, automatic generation of cross-platform user preferences and improved CnGAN (Perera and Zimmermann, 2019) for non-overlapped users. CnGAN is also used for recommendations.

2.2.2 Inspiration Stimulation

Research on the application of AI in creative stimulation can be divided into two areas: retrieval and the generation of design stimuli (Zhou et al., 2020). The former is mainly based on a data-driven approach, using different search algorithms to analyse existing databases to obtain stimulating information for inspiration through a process of classification, filtering, combination and analogy. The latter uses generative techniques to provide designers with creative stimuli by generating new stimuli. In terms of stimulus retrieval, Camera Obscurer (Singh et al., 2019) performs abstract image visual similarity searches on user-supplied photographs to find images that have inspirational value for graphic design projects. However, its data entries are rather limited, and its conceptualisation process is not fully understood. Combinator (Han et al., 2018), based on the theory of combinatorial creativity for image-inspired inspiration, helps non-professional and professional designers generate new inspirational images by combining related images in an overlapping format. However, as the images are only put together rather than combined, the result of such a combination is limited in terms of inspiration.

In terms of stimulus generation, the main focus is on visual image stimuli. The development of deep generative models has changed the nature of the creative process of AI-enabled graphic design to better assist in the generation of content for graphic design. For example, text-to-image (TTI) generation ((Xu et al., 2017)), automatic image colouring (Zhao et al., 2019) and image style migration (Gatys, Ecker and Bethge, 2015) are gradually being applied to creative stimulation in graphic design.(Han et al., 2018) provide a method for applying generative techniques to idea elicitation by retrieving relevant concepts through a semantic ideation network model and then building combinations of images of different concepts through generative adversarial networks (GANs). We found that there is still relatively little research in this area. At present, there is still a need to generate arbitrary combinations, and this requires large databases, significant training time and computational resources.

TTI is one of the most innovative areas of multimodal learning, aiming to automatically generate virtual images that meet the needs of users based on the information described in the text. Through machine learning or deep learning methods, TTI provides a new way of exploring, supporting and understanding certain art and creativity in areas such as image editing, video games, image art generation and computer-aided design.

Various companies provide AI visual tools based on pre-trained GANs and diffusion models to perform TTI generation. For example, somnai_dreams (dreams, n.d.),Disco Diffusion(aletts, 2021) and VQGAN+CLIP (Crowson et al., 2022), provide shared colab notebooks with customisable tuning parameters that can sample millions of images with detailed keyword descriptions or hints entered by the user. In addition, based on the diffusion model approach, they have numerous artistic tools dedicated to developing AI that can be manipulated quickly without code, style migration or image generation. Examples include (NightCafe Creator, n.d.), which can be sized to accommodate more keyword input.

In April 2022, OpenAI released DALL-E 2 (Ramesh et al., 2022), which can create original, realistic images and art from a text description. It can combine concepts, attributes and styles. With only a short text prompt, TTI models could generate complex images that never existed before while combining a semantic understanding of DALL-E 2 could modify existing images from a natural language caption, taking shadows, reflections and textures into account ((Ramesh et al., 2022). DALL-E 2 can also take one image and create several versions based on the original.

We found that while these generative models can inspire designers, they are often just images of unrelated elements combined, and although they do not look visually abrupt, they lack interpretability. The textual descriptions give greater scope for inspiring artistic creation. However, the steps of imagining, debugging, revising and adding to the process still require the involvement and control of the human creator.

2.2.3 Design Conceptualisation

Conceptualisation, the idea generation process of conceptualising a solution in the design process, is a creative process in which designers generate, develop and communicate ideas. The conceptualisation phase of AI-enabled design focuses on how to provide conceptual designs with varying degrees of fidelity, with a variety of conceptualisation options helping to produce more innovative design final solutions. CIP (Collaborative Ideation partner) (Kim, Maher and Siddiqui, 2021) using the design by analogy (DbA) approach To calculate visual and conceptual similarity metrics to match ideas based on sketches drawn by users, this study used two different components. One is a component based on vector representations of visual features of the sketches for similarity calculation. The other is based on the category names of the sketches and uses two pre-trained word2vec models to provide the designer with a high-fidelity diagram of the creative design. In this way, the user can freely sketch and edit in the rest of the drawing space, and the CIP will provide inspiring sketchs based on these two similarities. The results of the study showed that the AI-based stimuli produced different thinking outcomes compared to the random sketch stimuli and that the conceptual similarity metrics-based ideas were more novel and diverse.

Research into text description to idea sketch generation is also a hot topic. Sketchforme (Huang and Canny, 2019), an automatic sketch generation system based on the Sketch-RNN mode. and the GPT-3 model introduced by OpenAI, which enables the rapid generation of interface prototypes, can automatically generate sketches with corresponding semantics based on the user's text description. However, there is still little research on the application of text-based sketch generation, which has not yet been widely adopted, and in many design task scenarios, it is difficult to truly solve the problem of pattern innovation for design ideas from the perspective of user needs.

We found that the dataset remains one of the limiting factors for the richness and high quality of the output content of these research directions. The involvement of AI in the ideation process can effectively support the designer's creative process, but as hand-drawn sketches are inherently susceptible to influences on account of subjective factors (e.g. style), semantic text features act as constraints on the involvement of AI in the design ideation process and can present content with higher similarity and diversity.

2.2.4 Design Evaluation

Judging design content is a complex issue for design, as it can be subjective and 'fluid', and there is a wide range of styles that are difficult to choose from. In view of this, AI-led design evaluation aims to help users measure the outcome of a design so that it can provide them with relatively objective and effective advice. Related applied research in the field of visual design focuses on the aesthetic assessment of visual design, using computers that simulate the human mind, i.e. visual aesthetic beauty assessments, to determine whether or not a piece of work is beautiful or to what extent.

Since the development of Convolutional Neural Networks (CNN), these have been implemented for object classification and recognition and have made significant improvements in the accuracy of image aesthetic evaluation. (Lu et al., 2015) used dual-input deep convolutional neural networks with global and local views of images to automatically extract aesthetic features from images. (Krizhevsky, Sutskever and Hinton, 2012) built on their work by adding a saliency detection model to extract representative blocks of images. (Sheng et al., 2018) adaptively adjusted the weights of each image block based on an attention mechanism, and (Kong et al., 2016) combined image and image content information to classify images at an aesthetic level from 1–5. We found that although these methods do not require professionals to design the appropriate aesthetic features for different types of visual content, their prediction results still suffer from poor interpretability.

Applied research on aesthetics evaluation to support the design process is rare, focusing mainly on GUI (Computer Hope, 2019) and logo design. (Zen and Vanderdonckt, 2014) first used deep learning for GUI aesthetics computation. They used CNN for feature extraction, principal component analysis for feature dimensionality reduction and a support vector machine for classification. In addition, Webthetics was proposed by (Dou et al., 2019), who treated GUI aesthetic evaluation as a regression problem rather than a binary classification problem. To overcome the lack of web aesthetic evaluation data, they first used the Flickr style dataset (containing 80,000 images) and implemented model pre-training based on an image style recognition task, used a transfer learning approach to apply the image style recognition task to the evaluation model and then migrated the pre-trained model to the GUI aesthetic evaluation task. Zhang et al. (Zhang et al., 2017) conducted an aesthetic evaluation of LOGOs, where they quantitatively evaluated the balance, contrast, harmony and aesthetic value of each LOGO in the LOGO dataset, using a semi-supervised machine learning approach to construct a linear regression model for these manual ratings.

2.3 Status and trends of research on artificial intelligence-driven visual design generation

The application of AI as a major source of creativity in visual design is mainly represented by intelligent generation tools that address automated, batch design tasks. These tools focus on using AI techniques, including generative models, to help designers and ordinary users render creative ideas in the form of images by simply entering specific design goals and constraints, as well as to support the deepening and development of products. Most relevant research focuses on the automatic generation of layout generation, automatic colour matching, intelligent poster generation and many other areas.

Graphic layout specifies the position and size of visual typographic design components. Earlier work typically used templates, dynamic programming or some heuristic design rules to generate document layouts. In recent years, deep neural network-based generative models have found more use in layout generation because of their ability to learn complex, high-dimensional distributions. Common models such as LayoutVAE(Patil et al., 2020),uses the framework of variational autoencoders (VAEs), and LayoutGAN(Li et al., 2021),which uses the framework of Generative Adversarial Networks(GANs). For example, layoutGAN uses a GAN-based framework to model semantic and geometric relationships to generate layouts from Gaussian noise, laying out design elements in their proper place to form a Scene.In addition, it is possible to generate the framework and properties of each frame at the same time.Moreover, Neural Design Networks (Lee et al., 2020) offer a new approach for generating graphic designs from components based on user-developed attributes and conditional constraints, as well as a transformer-based layout model such as VTN (Arroyo, Postels and Tombari, 2021), LayoutTransformer (Gupta et al., 2021), BLT (Kong et al., 2022), which achieves better quality and diversity in layout generation than GAN or VAE models. However, these layout generation models focus more on graphical relationships and ignore other issues of visual design coherence, such as image content, font matching, colour choice, etc. In addition, these techniques do not focus on the real design, the designer’s perspective of a real design piece, and are accompanied by the problems of large amounts of design data and training complexity.

Colour is an important aspect of visual design and often influences the effect and expression of a design, with each colour palette conveying different feelings and meanings and evoking emotional resonance in the viewer. Users often need a certain level of experience in design theory and aesthetics, and by making repeated choices and adjustments to colours, they ultimately create visual harmony, balance and consistency. This poses a challenge for both practically efficient and untrained users. (Jahanian et al., 2013) and (Yang et al., 2016b) have developed different automatic colour design methods for image-text layout design based on colour harmony, colour semantics and these colour theories. In practical applications, there are many online colour design platforms (e.g. Adobe colour) that support the creation of colour swatches using colour wheels or images to give users colour-matching suggestions. (Cheng, Yang and Sheng, 2016) and (You et al., 2019) used training data to predict colour performance based on a data-driven approach by constructing large-scale colour image reference datasets. (You et al., 2019) proposed a method that could automatically colour match other constituent elements of an advertising image based on the input keywords and the colour of the main product element. The data-driven approach is considered more practical, but the quality and quantity of the annotated data have a significant impact on the final colour matching results. Also, colour harmony and colour contrast can appear to perform poorly in advertising images where the raw data lacks training images related to the input image or keywords.

In order to produce more complete and stylised graphic designs, there has been a lot of experimentation and research into the automatic generation of visual text typography. (Yang et al., 2016a) have designed a system that automatically generates digital magazine covers by summarising a set of topic-dependent templates (spatial layouts, semantic colours, harmonic colour models, and font emotion and size constraints) and introducing a computational framework that incorporates the key elements of layout design. (Guo et al., 2021) propose a system that displays posters generated based on the product image and taglines uploaded by the user, allowing the user to edit the posters by preference. However, users have to enter the title and content manually, and there is no link between text description and typography, which basically relies on the template setting.

3 ANALYSIS OF THE PRODUCTION MECHANISM OF TRADITIONAL VISUAL DESIGN CREATIVITY AND EXPLORATION OF INTELLIGENT USER NEEDS

3.1 Tools and methods for conducting user experience research

Qualitative and quantitative methods were used in this project to explore users' needs in the graphic design process. On the qualitative level, we used the observational method, verbal protocol analysis, in-depth interviews, focus groups, scenario analysis, user profiles and questionnaires. On a quantitative level, we use a variety of data analyses to attribute the collected data. We started from the user's needs, forcing the technology to transform and upgrade, and explored the opportunities for AI to intervene in the field of graphic design at the demand level, and established a stakeholder canvas based on the idea of User-Centred Design, summarising the three stakeholders: ordinary users, designers and software tools platforms, and The study also identifies three stakeholder groups, i. e. "professional designers" and "general users with design needs", and creates an accurate user profile. Our study includes a real data sample of 117 graphic designers and 17 graphic design companies/studios/team leaders, and a sample of 231 general users who have purchased design services or done creative design work, and 12 companies and branding/marketing business leaders.

3.2 Stakeholder map construction

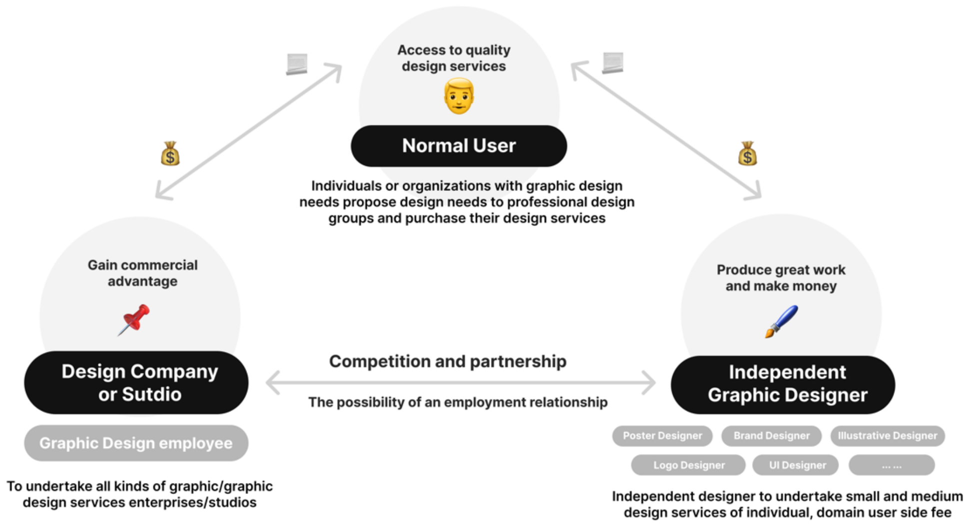


Figure 1: Graphic Design Industry Stakeholder Map

We first researched the current graphic design market size, design service models, design service types and market service prices. Then we conducted in-depth research on the service model, design process and market feedback of the visual design industry in different scenarios to obtain multiple data types and perform data analysis. The three main stakeholders of graphic designer services were identified: independent users with design needs (individuals or organisations), independent graphic designers and professional design services companies. The relationship and core interests of the three groups are shown in Figure 1. The core interest of the independent designer is to produce high-quality work and receive a financial return. The main core interest of the design company is to obtain commercial benefits. The core interest of the user with a design need is to get a high-quality design that meets their design needs in a short time and at the right cost. Through research, all three groups are involved in the graphic design process. We will summarise a user profile that is more relevant to this study based on the stakeholders, i.e. the group of professional designers and general users with design needs.

3.3 Constructing user journey map for graphic design creative process

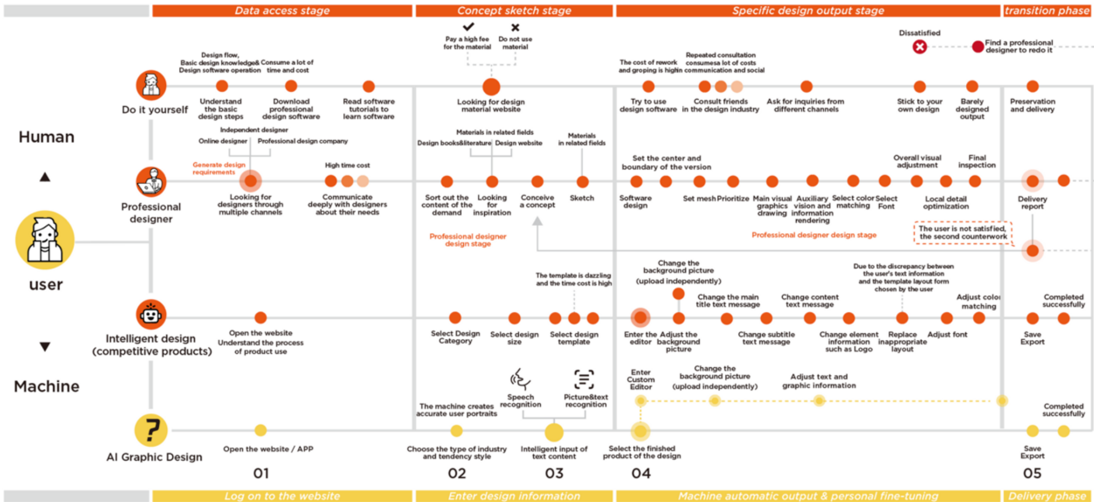


Figure 2: User journey map in graphic design process.

As shown in Figure 2, by studying the workflow of graphic designers and ordinary users in acquiring or performing graphic design, we extract similar high-frequency behavioural nodes in the process and provide a database for building a complete user journey map.

The above diagram shows the analysis of the behavioural nodes of ordinary users in accessing graphic design services. It is divided into four paths: ordinary users (1) learn professional design software to design on their own; (2) seek professional designers for graphic design services (including graphic designers' design workflow diagrams); (3) use current online template design platforms (e.g. Cannva, Gaoding); (4)An ideal AI collaborative graphic design process based on feedback from user research.

The user journey map shows behaviours between each different node of groups - professional designers and general users- in the graphic design process. It also exhausts the different graphic design production paths and the design processes of different design aided tools from a horizontal perspective. It provides a basis for later research to explore the design pain points under different personas and the opportunities for AI technology to intervene in the graphic design workflow.

3.4 User pain points and human–machine interaction opportunities

In the following, we will analyse the pain points of the graphic design process in three dimensions: graphic design industry, designers, and general users. And from these pain points derive points of opportunity in human-computer co-creation design.

3.4.1 Graphic design industry pain point analysis

1) Uneven business skills at the graphic design end

Following the research of 17 design enterprises/studios of different sizes and 117 graphic designers, it was found that the percentage of bachelor's degree holders in visual communication and related disciplines was about 45%, while the percentage of employees who learned graphic design through self-learning (training courses) was about 18%. Lastly, the percentage of employees with a professional design background but not a visual communication-related professional background was 36%. In summary, the proportion of interdisciplinary designers and half-way visual designers exceeded that of visual designers with a professional background by about 5%. As these designers are part of a training group, they have many problems in terms of professionalism, aesthetics, creativity and communication with clients.

According to a statistical report, by mid to late 2021,the current global design workforce is about 96 million (full-time) with 18 million (full-time) in China alone, and according to a report released by MGI (McKinsey & Company, 2019), the demand for jobs in China will grow to 85% in the next 30 years in the broad category of design and creative industries. Yet 85% of graphic design companies in China being small studios. Due to the low entry threshold in the graphic design industry, designers who have no prior experience in systematic design are common in small-scale design teams. More often, these designers are part of a trained halfway house group, with many problems in professionalism, creativity and user communication.

2) Inefficient overall collaboration in the graphic design workflow

The aftermath of the pain point of information and demand asymmetry is evident in the. overall inefficiency of this pain point, forming a vicious circle. The lengthening of the design cycle also leads to two outcomes: one is an increase in design cost, which is mainly the responsibility of the user, and the other is a decrease in design quality at the same price, which is mainly the responsibility of the designer. As such, an effective mechanism, platform or new productivity tool is needed to balance and mediate between the two.

3) Inefficient overall collaboration in the graphic design workflow

Always serving large corporations and brands, quality design resources pay little attention to the visual design needs of the public in their daily lives, with high prices and long lead times making it difficult for the average person to access and afford them.

The total global market for graphic design is currently over $820 billion, and there is a significant difference in demand for design between the high-end market and the mass market. The low spending power of mass market users has led to small and micro design agencies that serve the mass market having to increase their business in a short period of time to maintain their operations, and as the time allocated to each project is limited, this makes it difficult to guarantee design quality. In the human designer's game, good design and high cost are positively correlated, but due to the current state of the mass market's user spending power, it is rarely able to enjoy quality design services for a short period of time and at a low cost. Of the 231 users who have purchased visual design services, 80% felt that the final cost of the design delivered was high and that the results they received did not match the cost of the design they paid for. In addition, 45% of users had to use 'free designs' provided by the producer or make their own 'design patchwork' using simple tools owing to the high design costs.

The data also shows that rough and low-quality graphic design has now taken over people's everyday environments. Around 65% of the 231 general users surveyed said that visual graphics in urban environments are not aesthetically pleasing and that they are not satisfied with the visual environment in their neighbourhood. There is a tendency for people to adapt to the 'wild design' that pervades their everyday environment after a long period of low-quality visual impact. Especially in China's non-first-tier cities, these visual elements surround the public in an almost all-encompassing way, so to speak, and 'they' have become part of the urban 'landscape' in which people are passively accepted or influenced, creating a sense of 'This is the way things should be'. The result is that the evaluation of 'beauty' has been greatly influenced, leading to the struggle in China in recent years for basic aesthetic education. We believe that the root causes of this phenomenon are twofold. On the one hand, the subjects who communicate their messages for commercial gain or to export their values do not have enough money to specialise in the visual expression of their 'brand', often using the producer's crude templates and the most conspicuous fonts or colours to highlight their products. On the other hand, these messengers do not have a sense of brand building. With the trend of consumer upgrading, there will be an increasing demand for uniform and coordinated high-quality, differentiated brand visuals.

3.4.2 **Designer's pain point analysis in graphic design creative process**

We conducted a questionnaire survey of 117 frontline graphic designers and 17 design company directors, including the following indicators: design industry, job content and type; time investment ratio in different design stages; average daily working hours; average daily output of visual design solutions; work completion ratio; common software; common materials/case websites; design originality; work happiness/stress value and monthly salary. Our analysis has led to the following two core pain points:

1) **In view of the high intensity of the graphic design industry, it is difficult for designers to maintain intensive creative inspiration even for a short period of time. At present, designers generally adopt the single approach of 'design case sharing websites' and 'design material websites'. This approach may lead to the homogenisation of design results and a reduction in the designer's originality, which is one of the common pain points at present. Therefore, the use of AI technology to stimulate designers' creative inspiration and visualise concept designs quickly provides an opportunity to alleviate these pain points.**

According to the data feedback, 80% of designers believe that the creation of poster design entails finding the right intention and design materials to start, while 75% of designers need to spend more than 45% of the overall design process in finding and collecting materials. At the same time, 60% of designers are less satisfied with the current design materials on the Internet. Their satisfaction is distributed as follows: 55% visual style matching, 20% material resolution matching and 25% copyright costs.

The data also shows that while 35% of designers choose to create original design material, 65% often choose to purchase commercially available design material for fine-tuning to be able to complete intense design tasks quickly. Although 85% of designers have heard of online template design platforms like Canva and the Lupin AI design platform, 30% try to use this type of platform, and only 5% say they do design on this type of platform regularly. As much as 90% of designers say they have been asked to produce five or more visual design proposals in a day, and 60% of professional designers are forced to work overtime (upwards of 8 hours working time per day) more than three working days in a week. Of these designers, 70% feel that having a constant stream of creative design proposals in such short time periods places enormous pressure on them. In addition, as the output is not their own approved design ideas, their sense of achievement and happiness is low.

In summary, as it is difficult to maintain abundant creative inspiration in the overall graphic design workflow as well as in the early conceptual design stage, designers need to spend a lot of time finding the design intention style and suitable available design materials, tasks which make up a great deal of the time and costs in the overall design process. This results in a low input-output ratio (ROI) per unit of time for designers. Under the objective conditions of constant task intensity, most designers work longer hours and fall into a vicious cycle of work accomplishment and low happiness.

2) **Due to the low level of intelligence and the cumbersome operation of current design software/tools, designers generally believe that they spend a lot of time on repetitive labour tasks in design, such as the tedious process of using current mainstream design software tools to realise conceptual designs, while investing little time in the design ideas themselves, which is currently a common pain point for designers. As such, using AI to reduce the repetitive workload of graphic design for practitioners in the operation of tools and software is one of the research opportunity points. Due to the low level of intelligence and the cumbersome operation of current design software/tools, designers generally believe that they spend a lot of time on repetitive labour tasks in design, such as the tedious process of using current mainstream design software tools to realise conceptual designs, while investing little time in the design ideas themselves, which is currently a common pain point for designers. As such, using AI to reduce the repetitive workload of graphic design for practitioners in the operation of tools and software is one of the research opportunities points.**

   Current graphic design tools are mainly based on a single subjective operation of the designer. While this can only realise the process of digital visualisation of graphic design solutions on the basis of a clear design intention, it lacks the support of early creative stimulation driven by intelligence and the optimisation and iteration of human-computer collaboration in the middle and later stages of design solutions. Around 75% of designers think that using various methods to realise the conceptual design on design software (e.g. alignment, bevelling, cutting, multiple layouts, colour matching) is very tedious and time-consuming, and 95% say they spend very little time on creative work.

3.4.3 **General users' pain point analysis in designer graphic design creative process**

We conducted a survey on 231 general users who had purchased graphic design services or carried out graphic design. The following indicators were included in the survey: occupation and industry; type of graphic design needs; frequency/month of graphic design needs; delivery cycle of single design service; involvement in the design process; frequency and proficiency of using design platforms/tools, matching design expectations; and satisfaction with design results. The questionnaires and in-depth interviews were conducted across different dimensions, such as the cost of single poster designs and the proportion of designers using online design platforms and intelligent design tools. Our analysis has led to the following two core pain points:

1) The current mainstream graphic design software has a high threshold for use, and the high learning cost for ordinary users without design skills is one of the common pain points at present. While in the current design service market, some parties with design demands generally come with clear design needs and ask professional designers for skilful concept visualisation, the lack of creative visualisation tools that can support non-professional designers in a rapid manner is one of the common pain points at present. As such, knowing how to use AI technology to optimise existing design tools and to generate graphic designs through natural language descriptions for ordinary users is one of the current research opportunities.

2) The design demand side (ordinary users) generally perceives information asymmetry and demand mismatch with the design supply side and low matching of design results with expectations, leading to repeated revisions of design solutions back and forth, resulting in **inefficient overall design services is a common pain point at present. Therefore**, **how to use AI to coordinate the consistency of language perception between the design demand side and the design supply side (machines) is a current research opportunity point.**

   The questionnaire and interview survey results were collected from a sample size of 231 general users and 117 designers from different industries who have purchased design services. Among the group of designers, 70% of them said they had experienced communication problems with the A-side, involving non-repetitive communication requirements, inconsistent style understanding and differences in aesthetic tendencies, resulting in ineffective revisions of design solutions and wasted time. The data also shows that among the general user community, 60% consider the designer's first output to be unacceptable and in need of a second design, 35% consider it acceptable but needing to be optimised and changed and only 5% of users could accept the designer's first design output. Visual designers and users are prone to friction at the level of communication, with designers believing that users often show a strong sense of control during the communication process and that their unreasonable demands make it difficult to control the quality of the design work.

   This explains why most design projects are carried out without effective communication between the user and the designer. The information is not communicated properly, which ultimately leads to a mismatch between the needs of both parties. With the involvement of AI in the graphic design process, we are also faced with the same incommensurability in input language descriptions between humans and machines. Accordingly, how to use AI to coordinate the cognitive congruence of language descriptions between the design demand side and the design supply side (machines) is the focus of this research.

4 RESEARCH QUESTIONS AND OBJECTIVES

Based on the literature review and in-depth user research above, we have identified a diversity and richness of layouts and important trends in the development of AI technology in the graphic design field, which is increasingly becoming the 'raw material' of the toolbox of creatives and design practitioners. It helps them visualise faster and in potentially impossible ways and changes the perception of how we might create specific types of creative assets in design. We, therefore, set out the core problem statement and the intended research objectives for this study.

4.1 Core research questions

1) The lack of user experience-centred AI graphic design research has been caused by the emphasis on technology and the lack of demand. The problems of concept design inspiration and design efficiency are common pain points in the graphic design process.

2) The current design software or tools are not yet intelligent and cumbersome. Designers generally feel they spend a lot of time on repetitive tasks, for example, when they use mainstream design software tools with cumbersome processes to realise concept designs, a common pain point for designers today. Therefore, one of the research opportunity points is how to use artificial intelligence to reduce the repetitive work of graphic design practitioners in the operation of design tools and software.

3) The current mainstream graphic design software has a high threshold for use and is costly to learn for the average user without design skills. And in the current design service market, some design demanders will come with clear design needs to hire professional designers for graphic design. The lack of visualisation tools that can support non-professional designers to visualise visual ideas quickly is a common pain point. Therefore, one of the current research opportunity points is how to use AI technology to optimise existing design tools and realise the process of generating graphic designs through natural language descriptions for ordinary users.

4) There is a general perception of information asymmetry and demand mismatch between the design demand side (general users) and the design supply side (designers). The situation leads to a low match between design results and expectations, resulting in repeated modifications to the design solution and causing inefficiencies in the overall design service. Therefore, how to use AI to coordinate the consistency of linguistic perceptions between the design demand side (users) and the design supply side (machines) is a current research opportunity point.

4.2 Expected objectives of the study

In order to determine which design phase to focus our research on, we conducted a deep dive into user requirements through a stereotypical combined with a quantitative research approach to existing AI technologies, including NLP, deep learning generative models, multi-modal neural networks, machine learning algorithms, and other fields. We focused on the use of large-scale pre-trained models to provide design paths and design content quickly and creatively for the creative inspiration and ideation stages of graphic design. A new feasibility of combining AI with graphic design is developed, and therefore the focus of this thesis is on how to apply AI as a 'raw material' to a more logical and implementable design conceptualisation process, resulting in a reusable AI design approach that can address real needs. In summary, in this research, we explore a combinatorial generative approach that combines large-scale training models as the underlying tool with real-world needs in design. The actual needs of users in graphic design are studied from a user-centered perspective, and new perspectives and ways of AI-enabled graphic design are explored.

The expected objectives of this study are as follows:

1) In the case of e-commerce advertising, an innovative generative framework with unstructured input is built based on a combinatorial generation approach, using a large pre-trained model as the underlying tool. The result is a framework that allows users to generate a variety of sketches of poster ideas with any natural language input. The framework integrates multimodal information, such as style, text and images, by extracting keywords to generate associations.

2) Build a user-centric AI-generated creative inspiration sketch prototype to support the interactive process of automatically generating typography.

3) The effectiveness of the solutions was evaluated by means of a survey on the evaluation of these solutions by non-specialist designers and professional designers.

5 RESEARCH CONTENT AND PROCESS

In this study on NLP, deep learning generative models, multi-modal neural networks and machine learning algorithms were involved.

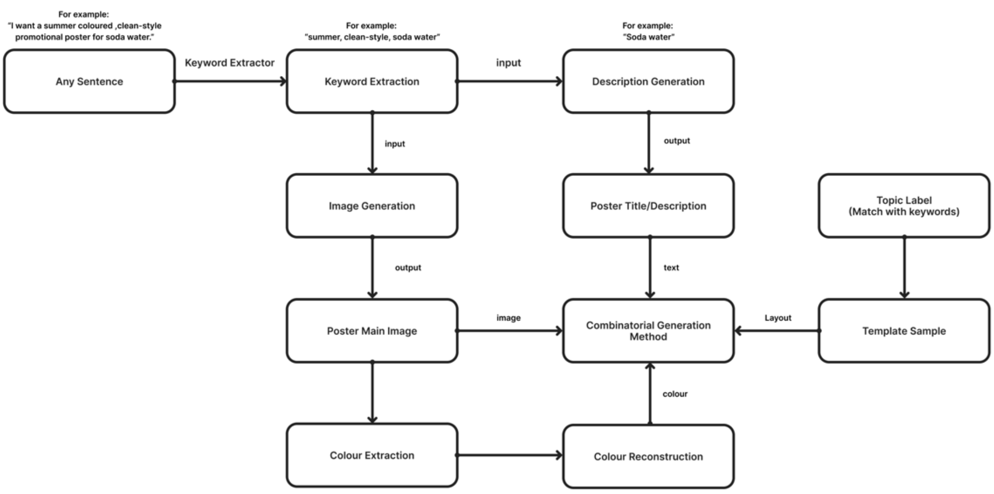


Figure 3: Overview of the workflow of the Combinatorial Generation Framework

Based on the analysis and summary of previous work, we explored a combinatorial development framework for automatically generating visual sketches with user preferences from natural language. Figure 3 illustrates our expected workflow. The framework consists of six main modules: (1) keyword extraction from the natural language entered by the user to determine the poster theme; (2) generating the product descriptions in the poster from the keywords; (3)

generating the main visual images; (4) reconstructing the colour scheme based on the generated images; (5) developing the layout of the theme template; and (6) combining them to generate sketches of the combined graphic and text layout. In the following, we will further explain how we have implemented each module.

5.1 Keyword Extraction

We chose a keyword extraction framework known as TextRank(Mihalcea and Tarau, 2004) in order to identify the main information in the natural language entered by the user and to associate it with multimodal information, such as sketch style, text and images. In an unsupervised manner, it first constructs a network from the adjacency of words, then iteratively computes the importance score of each word in a long sentence using a PageRank-like algorithm and then ranks the importance scores to obtain the keywords. In this study, we chose TextRank4ZH for Chinese text.

Based on the preliminary research on the whole process of graphic design, we constructed a high-frequency natural language description structure for users to obtain design results, and the specific data sources are located in the following two dimensions: first, the analysis of high-frequency keywords input by users when looking for design templates and design elements in the design process, and second, the analysis of high-frequency design demand keywords provided by the designer in the design service process. Through the analysis, we extracted the keywords style description, application scene description, subject description, colour description and size format description, and established a common 'input content' that is in line with the conventional user input habits, can be quickly understood by the machine and generated with good results. The keywords are 'description of the scene, description of the subject, description of the colour and description of the size and format.' Here we take '**I want a summer coloured, clean-style promotional poster for soda water.**' as an example to test the keyword extraction and to use it as a prompt for each of the modules in the paper.

TextRank4ZH first segments the given input according to complete sentences. Next, the sentence is split and annotated with this line, and the stopwords are filtered out. In order not to be distracted by names specific to e-commerce platforms, etc., we add a series of these words to TextRank4ZH's original stopwords, such as 'poster', 'department', 'publicity', and 'advertising.' By removing these stopwords, TextRank4ZH retains only words of the specified lexical nature, such as nouns, verbs and adjectives. Also, some proprietary design terms were added to the dictionary, such as 'vintage style', 'fashion style' or 'clean style' in order to extract the design style precisely. The algorithm then calculates the most important words and marks them in the original text. If adjacent words are formed, they are combined into multi-word keywords, and finally, we get the extracted keywords as '**summer, clean-style, soda water**' (Table 1).

Table 1: Example of TextRank4ZH.

original sentence	我想要一张夏日色系的，纯净风格的苏打水的宣传海报 (I want a summer coloured ,clean-style promotional poster for soda water)
words_no_stop_words	一张，夏日，色系，纯净，风格，苏打水 (one, summer, colour, clean, style, soda water)
words_no_dict	夏日，风格，苏打水 (summer, style, soda water)
words_all_filters	夏日，纯净风格，苏打水 (summer, clean-style, soda water)

5.2 Description Generation

5.2.1 Data pre-processing

In the text description generation stage, we used the CEPSUM2.0 dataset, a Chinese e-commerce product summarisation dataset containing approximately 1.4 million product-summary pairs, separated into three product categories, including home appliances, clothing, and cases and bags. However, the product categories in this dataset do not cover the full range of categories on existing e-commerce platforms.

In order to improve the richness and completeness of the dataset, we carried out pre-processing work on the original data volume. First, we organised the dataset into four dimensions based on the model's data requirements: product category, product keyword, product title and product description, and then we delineated 16 main product categories based on e-commerce product categories, namely 'Digital Home Appliances', 'Stationery & Office', 'Clothing & Shoes', 'Home Life', 'Books & Entertainment', 'Beauty & Personal Care', 'Pet Supplies', 'Outdoor & Fitness', etc. The 16 main product categories were then broken down into specific product keywords, including 'mobile phone', 'computer', 'Headphones', 'Electronic audio', etc. Some of the categories were selected by searching for product keywords in the original dataset. For the rest of the data, we initially tried to use crawlers to collect the data, but due to the lack of download resources and the complexity of anti-crawler protection, we manually collected the product descriptions under each product keyword from the internet. We used the data from the product keyword description pairs to train our NLP model. The 'product titles', e.g. 'Smart and convenient mobile phone', are used as the 'title layer' in the subsequent combination generation module.

5.2.2 Generate corpus by Gpt2, Bart, Pegasus and T5

Textual content in graphic design not only provides more detailed information about the design, such as product names and descriptions but is also an important component to enhance visual communication, improve the appeal of the work and give aesthetic value to the design layout. In order to get a better text generation effect, we fine-tuned four Chinese pre-training language models in this module, BART, T5, Pegasus (huggingface.co, n.d.) and GPT2(logCong, 2021) respectively. The dataset, which is prepared in sets of product keywords-product description, was divided into two sets for training and testing, each being 80% and 20% of the prospective whole. The training set contains 226,634 characters, and the testing set contains 56,145 characters. The four models were trained for 10, 20 and 50 epochs and the final generation process, and results were evaluated qualitatively and quantitatively for comparative analysis. The three models, BART, T5 and Pegasus(huggingface.co, n.d.), were integrated, and finally the same run.py and generate.py files could be called for training, generation and fine-tuning of the models.

The results are shown in Table 2 and Table 3.

Table 2: Experimental results for models when using the full training.

Model	R-1	R-2	R-L	BLEU
BART-139M	7.96	3.14	7.45	5.85
Pegasus-238M	7.82	2.88	7.34	5.25
T5-77M	7.34	3.13	7.39	9.54
GPT2-NewsTitle	3.62	4.13	3.28	4.70

Table 3: Several examples of using the trained model to generate product descriptions, where we set all the input-product title to "soda water"

Model	Product description
BART-139M	澄净天然果香，品质有保障 (Serene natural fruit, quality guaranteed)
Pegasus-238M	气泡更加丰富细腻，口感醇厚，温柔不乏个性 (Bubble richer and more exquisite, taste mellow, gentle personality)
T5-77M	清新自然 (Pure and fresh and natural)
GPT2-NewsTitle	采用优质环保塑材制作，环保无味安全 (Made of high-quality environmental protection plastic material tasteless safe environmental protection)

From the table, we can see that the performance of GPT2-NewsTitle is significantly lower than that of the other three models, and the fact that the given exams are not in the right text also indicates that the model does not work well in this scenario. The reason for this may be that the dataset does not exactly match the scenario, resulting in a difference between the generated text fields and the commodity description generation task. GPT2 is a generative pre-trained machine learning model created by OpenAI, whose basic aim is to predict what words will appear next after some hints of seed text. The model was trained on over 40 GB of internet text. For this project, the model we chose to fine-tune is a Chinese news headline generation task, GPT2-NewsTitle, based on HuggingFace's transformers, which belongs to the Text Summary class of NLG(Natural Language Generation) tasks, generating concise and fluent news headlines while retaining the basic information and overall meaning. However, for this project, we have tested that it is more appropriate to use the encoder-decoder model to do the generation rather than the one-way decoder model.

For the BART, Pegasus and T5 models, there is little difference in performance, with BART having the best relative performance and Pegasus the second best; however, T5 is higher in terms of BiLingual Evaluation Understudy (BLEU) metrics. It can therefore be directly fine-tuned for generative tasks. Pegasus is also a good contemporaneous model for text generation tasks; however, it is not tuned for Chinese, while T5 supports multiple languages and has relatively good support for Chinese, which may be one of the reasons for its higher BLEU score. BiLingual Evaluation Understudy can indirectly respond to a certain degree of uniqueness for generative tasks, as BLEU goes to smaller values for the number of occurrences of a phrase in ref and hyp. So if a phrase is repeated too many times, it will not score as high. But after all, BLEU evaluates words and cannot assess semantic features, such as coherence.

We also use the Recall-Oriented Understudy for Gisting Evaluation (ROUGE) metric as an evaluation metric for our model, mainly based on recall calculations. Where ROUGE-N mainly counts the recall on N-grams, ROUGE-L considers the longest common subsequence between the generated results and the original data. As can be seen from Table 2, the results of the GPT2 model vary considerably. Although ROUGE was used as our primary evaluation to measure the quality of the generated commodity descriptions in direct similarity to the original commodity description data, this may not be a reasonable assumption. This is because there is little overlap between the valid descriptions generated by the model and each other. Therefore, the best evaluation will be performed by humans.

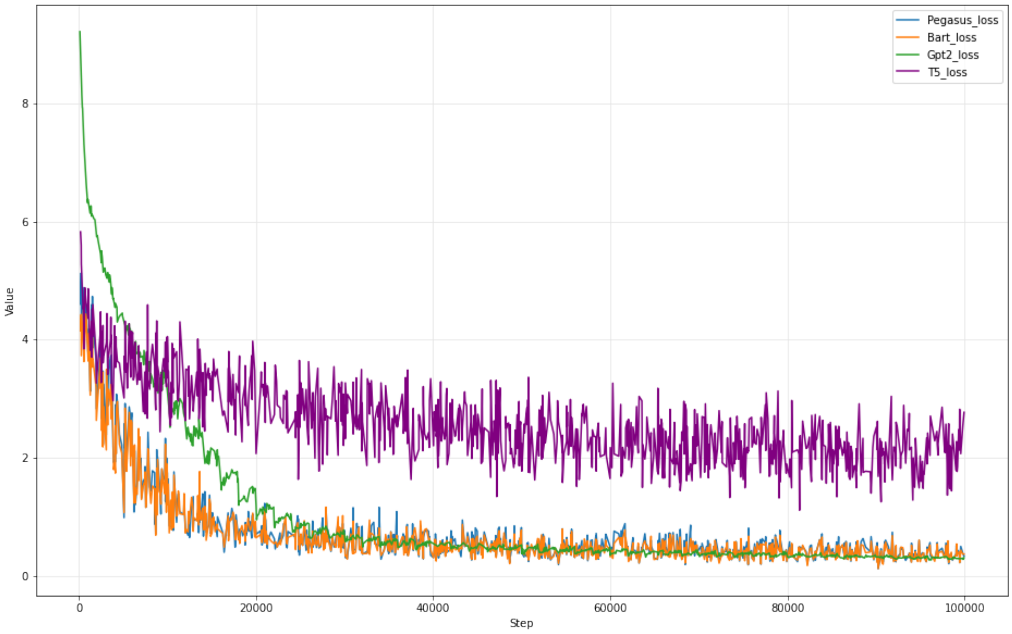
We also note that for the BART, Pegasus and T5 models, the descriptions generated by the models for the specified items are fluent and semantically appropriate. However, the descriptions were fixed and lacked variety, and even changing the random seed and adjusting the number of beams did not improve them significantly. There is a link between this phenomenon and the characteristics of the task, and the following is a possible analysis of the causes:

1) The limited variety of commodities and the high level of redundancy (e.g. black and green teas are similar), many of which are already included in the dataset, limits the output space of the model to a small range, thus affecting diversity.

2) The short length of the input product title provides insufficient information and limited semantically relevant product descriptions.

More perspicacity concerning the training of four models can be gained by looking at loss graphs (Table 3). It can be seen in those figures that loss gradually decreases if we train a higher number of epochs (iterations). It is observed that the losses all start to fall rapidly in the early training period and then continue falling slowly. This indicates that the model is converging and there is no significant overfitting. BART has a higher accuracy rate than the other methods, performs slightly better than other models, and is more computationally efficient.

Table 3: Loss graphs for each model after 50 epochs of training



In this module, we demonstrate that it is fundamentally feasible to generate product descriptions using the pre-training language model, but more methods are needed to optimise them, for example:

1) Further adjustment of model parameters

2) Improving the quality of the dataset corpus

3) Adding additional input information to increase the diversity of model outputs

At this stage, we are able to use the product keywords extracted in the keyword extraction module as the final text input for the 'subtitle layer' of the automatic portfolio generation module.

5.3 Image Generation

In order to solve the problem of image design material for users in the poster design process, we evaluated existing TTI models and found that DALL-E 2 is significantly more expressive than adjectives for nouns in prompts. It can recreate a high level of semantic understanding of the subject words but slightly lacks in detail and artistic rendering. Compared to the stable diffusion model, it is more suitable for generating realistic images rather than images with a strong artistic atmosphere. Based on the fact that this project is based on an e-commerce platform, we decided to use DALL-E 2 as the underlying tool for the image generation module of this project.

In the testing phase, we tried to directly use the original sentence, "I want a summer coloured, clean-style promotional poster for soda water" as the prompt for DALL·E 2 to generate a picture. The result is shown in Figure 4. As shown, the prompt cannot make DALL·E 2 work expectably. Although the tone of the picture and the main visual picture are good, there are garbled characters and dislocations in the text part, as well as blurred edges. The model does not accurately encode the spelling information of the rendered text, as if it is not "creating" as we want. This is a common problem of all generative models right now. For objects with specific semantic details, it is often impossible to produce better results. The specific semantic details here, intuitively refer to textual details, and abstractly refer to details such as quantity, location, and relationship. We believe that because these details are high-frequency components, and since high-frequency components account for a small proportion in the loss function, deep learning tends to preferentially use low-frequency components to fit the objective function. We call this mechanism the F-Principle (Xu et al., 2020). At the same time, because the BPE encoding used by DALL·E 2 masks the spelling of words in the title for the model, and the model needs to see each token written in the training image independently, it can better learn to render the text.

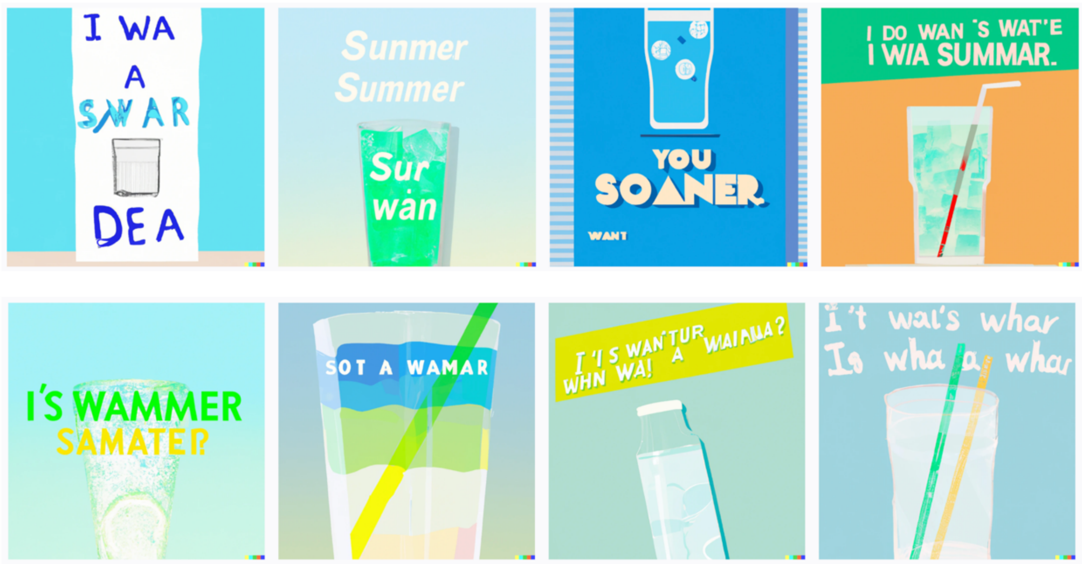


Figure 4: Using the keywords extracted by the keyword extraction module as the input prompt in the image generation module.

Therefore, in this project, according to the characteristics of DALL·E 2, we choose to input prompts that are more "suitable" for DALL·E 2 by extracting keywords. Let's take the aforementioned "summer, clean-style, soda water" as an example, DALL·E 2 can combine concepts, attributes, and styles (Ramesh et al., 2022a), through our input text it could generate complex images that never existed before and combine a semantic understanding of relevant and irrelevant objects. As shown in Figure 5, these images can be based on the primary visual sources that generate the final poster idea sketches. From the generated results, DALLE2 presents quite good photo-realism while maintaining more variety.

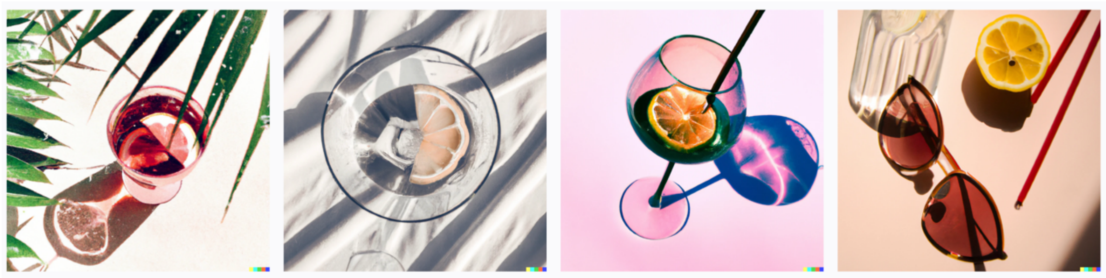


Figure 5: Example of the image generated from the prompt we entered.

At the same time, because of the privacy settings of dalle2, the prompts that are different from stable diffusion require artist style to produce amazing pictures, but there are also situations where the styles of related artists overlap too much. After testing, DALL·E 2 will not “plagiarize" artist style's style of painting based on input prompt. For humans, DALL·E 2 can be considered more of a "creator".

5.4 Colour extraction and reconstruction

Colour reconstruction in graphic design involves recolouring the remaining design elements according to the colour of the main component while maintaining the colour harmony of these elements with the main element after the modification. In this project, in order to achieve a more harmonious and semantic colour reconstruction system, we take the output of the image generation module as the input information of the colour reconstruction module. First, we use Extcolors (CairX, 2020) to quantify and visualize the colours of the generated images. Then, we use the rules based on the colour wheel palette to define the corresponding colour design rules according to different theme templates. For example, complementary colours, analogous colours. Therefore, when our main visual image is replaced, the text and background colours will be transformed according to the defined rules to achieve intelligent colour matching output.

In this module task, first we use Extcolors, which colours are grouped based on visual similarities using the CIE76 formula (CairX, 2020), and rgb2hex library. Extcolors returns RGB values, which will be converted to HEX colour codes using rgb2hex. We set the tolerance value to 12 and limit the number of colour codes output to 3 colours (limit=4), then use the rgb2hex library to define a function to convert the RGB codes to HEX colour codes and create a dataFrame. In the end, we extract the the representative colours are visualized as shown in Figure 6:

Figure 6: Example of a visual representation of the quantized colours of a picture generated by DALL·E2 using extcolours and rgb2hex.

The colour wheel is a visual representation of the colours found in a prism, arranged in a circle, with the primary colours (yellow, red, and blue) spaced evenly around.Artists of all kinds—painters, quilt makers, web designers, graphic designers, interior designers, etc—use it as a basis for working with hues, shades, and colours. It's a great tool to plan colour schemes and colour mixes.

The colour wheel design rules with an algorithm is shown in Figure 7:

```
r, g, b = ImageColor.getrgb(main_color)
    h, l, s = colorsys.rgb_to_hls(r/256., g/256., b/256.)
    h += color_patch[0] / 360.
    if color_patch[1] >= 0: s = color_patch[1] / 100.
    if color_patch[2] >= 0: l = color_patch[2] / 100.
    h -= int(h)
    if h < 0: h += 1
    return '#' + ''.join(['%02x' % round(v * 256.) for v in colorsys.hls_to_rgb(h, l, s)])
```

Figure 7: The main code representation of the colour wheel rule.

The basic logic of the algorithm is: given a mainColour, and a colourPatch, such as complementary colour, analogous colour, etc. The corresponding new colour is thus calculated. In detail, we first convert RGB to HLS, and then add H to the specified angle dH, which means to rotate dH degrees counterclockwise on the colour wheel; if new L and S are set, then update. Its then converted to RGB representation and output as hex colour value, where param "main_color" gives the hex representation of the colour. The param "color_patch" represents the hue difference, a triple (dH, S, L), the range of dH is 0~360, which represents the counterclockwise rotation angle of the colour on the colour wheel. S (saturation) and L (lightness) range from 0 to 100, representing the saturation and lightness of the target colour, respectively. In the end, we formulated different rule definitions for the 6 theme models based on the style tones. For example, for the "nature" style template, we set the rules to not change the background colour, and then used the lighter colour in the image extraction colour # FBF9F5, for the text part, we define the principle of complementary colour, add colorParch, dh=160, S=70, L=70 to #003228 in the extracted colour, and finally get our text colour #e97db, the final effect is shown in Figure 8:

Figure 8: Apply the colour wheel rules based on the generated image, the effect of the "nature" template after colour reconstruction.

However, due to the randomness of image generation, the local colour of the image may be too light or too dark. Consider the global colour coordination and local readability, we should further analyze and adjust the special situation.

5.5 Theme template creation

It is a challenge to combine the multimodal information we have, and build a rendering engine that can automatically generates poster sketches -- how to lay out the relationship between the attributes of each design element will directly affects information perception and aesthetic experience. Although it was found in our previous research work that LayoutGAN, VTN, ContentGAN, etc. have been able to automatically laying out design elements into proper positions to form a scene, the amount of data and workload to train and to generate layout are large, and its final effect is uncontrollable, which is not the focus of this project.

The test method we proposed was inspired by previous works, and we investigated numbers of graphic design posters from different design platforms. and finally scratched typesetting designs from Pinterest from two dimensions: basic visual layout and emotional style classification. Based on those two dimensions, six topic templates have been temporarily developed, namely "vintage", "technology", "cute", "clean", "fashion" and "nature" as style labels, which have six different layouts, such as "Left and right", "Symmetry", "Centre", "Bespread", etc. Style labels are used to enable the extracted style keywords to match the template.

But we found that in the case of limited existing design information, it is difficult to distinguish different design styles from different typography alone. We tried to match the colour in the colour reconstruction module with the style label, adjust the colour extracted from the image to the colour value that matches the corresponding style tone, and then apply it to the text and background colour. In addition, according to the results of our user research, the proper selection of fonts is a common pain point at present. At the same time, fonts, as a visual element, will stimulate a certain perception of people, and it is also a very important factor affecting the overall visual design. Therefore, we linked font sentiment to style tags, and for each style tag matching template, we pre-set 2-3 fonts with similar sentiment perceptions, for example, the template for the "vintage" tag we set draws from 20th century Chinese font "XiQueJuZhenTi" imitating the Song Dynasty style, and the "HYChangMeiHeiJ" combining Hei and Song style, these two serif fonts make people feel nostalgic and Chinese style.

After researching various image formats, we decided to use PSD as the template format and PNG as the output image format.(Figure 9) This is because PSD contains layer information, which makes it easier for designers to create and modify templates using design software, and is also readable by the program, making it significantly better than other formats in this scenario. As a universal image format, PNG supports transparent channels, allowing the generated poster images to be displayed on various devices without any differences.

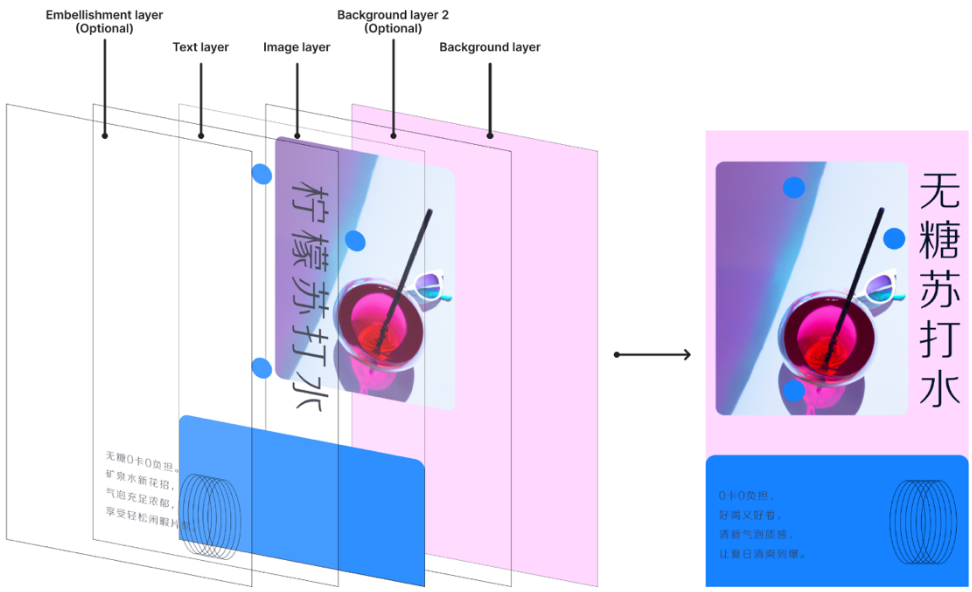


Figure 9: A Schematic representation of the template structure, consisting of "background layer", "image layer", "text layer", and "embellishment layer".

The procedure requires that the naming and type of layers in the template adhere to basic conventions：

1) The product name and short description text layers should be named “title layer/subtitle layer”, respectively, and should be of the type ‘Text Layer’ to allow the program to read information such as line numbers and fonts.

2) The main visual image layer of the product is named the “image layer”; no specific type is required, and masks are allowed.

3) The background layer is named the “background layer”. There is no special requirement for the type, and it should contain only one colour.

4) The embellishment layer is named the “embellishment layer”. There is no special requirement for the type; it should contain only one colour, allowing deletions and changes.

5.6 Combinatorial generation engine

Using an upstream AI generation model, we extracted keywords from the user input text and further associated them to generate relevant multimodal information, such as style, copy and images. The combined generation engine automatically matches the input style tags with pre-set templates and organically integrates the aforementioned multimodal information to generate posters that can be used for commercial promotion.

The output of the combined generation method is the generated poster sketch, the input of which includes the following information:

1) Theme templates

2) Style labels (for matching templates and calculating colourways)

3) Product titles and short product descriptions and their fonts

4) Main visual image of the product

The engine loads all the templates when it starts, forming a mapping between the style tags and the templates. In order to generate a poster image, the program first matches the most suitable template according to the style label, parses the layer information provided by the template, writes the product name, product description and main visual image into the relevant layers in the original layer order and adjusts the colour of some elements according to the calculated colour scheme. After completing the above steps, the program will overlay all layers in order to produce the final poster image.

The engine has the following main implementation methods and processes:

1) The engine is implemented in Python 3 and uses the PSD tools of third-party libraries to read the layer information from the PSD theme template file, perform the rendering process using PIL and output it as a PNG image.

2) The program parses each type of layer in turn in the order of the layers, extracting mainly the following information:

   a) Text layers: position, size, available fonts, number of lines, alignment, vertical or not, etc.

   b) Image layers: position, size, mask position and size (may be empty)

   c) Background layer: position, size, initial colour.

   d) Embellishment layer: position, size, initial colour.

3) When processing the rendering task, the program renders each layer in turn as an RGBA image in the order of the layers, taking into account the input text, image, colour scheme, etc., superimposed on the existing resulting image, and finally saving the result to a local file.

During the testing of the engine, a number of problems were encountered and improvement solutions were adopted:

1) **The input text and the text in the template may be different in length and number of lines, making direct replacement difficult.**

   a) Determine the number of lines. The program splits the input text with a line feed (\r or \n) and, in the case of multi-line text, uses the number of lines of that text directly; otherwise, the program cuts the input text into the same number of lines in proportion to the number of lines of text in each line of the template text, in order to keep the look and feel more or less consistent.

   b) Determining line height. The program looks for the line with the most words in the multi-line text and estimates the width of that line, which in turn determines the aspect ratio of that line. As the final length of the line should be equal to the original text width of the template, the height of the line can be calculated. Other lines of text will not have more words than this line, so the same line height is used.

c) Determine the line spacing. We already know the total height of the original text of the template H, the line height of the input text h and the number of lines n. It is easy to calculate the line spacing as $h' = \frac{1}{n-1}(H - nh)$.

d) The offset and spacing of each line are calculated based on attributes such as alignment and whether the line is vertical. By combining the kerning information of the font with the previously calculated data, the absolute position of each character in the resulting image can be obtained. Based on the above strategy, the program can render the replaced text to match the template.

2) Rendering of vertical text.

For vertical text, the reading direction is right-to-left and top-to-bottom, rotated 90 degrees clockwise relative to the regular horizontal text, so we can reuse the rendering logic for most horizontal text to rotate the result 90 degrees counterclockwise. During the rendering process, care needs to be taken to adjust details, such as word spacing for different types of characters.

3) The size of the input product's main visual image and the template's pre-defined image may be different and need to be adaptively resized.

Our goal is that the replacement image does not cause blank areas in the resulting poster and that the original image's aspect ratio is maintained. Making the image be cropped in as few parts as possible while satisfying the previous constraints. Specifically, a zoom-in operation is required if the input image is smaller than the template image. Otherwise, a blank area will appear; a zoom-out operation is required if the input image is larger than the template image. Otherwise, the program will crop an excessive area. The program will zoom in to cover the template's original image, cropping out the parts outside the specified area.

At the same time, we also found some limitations during testing:

1) When the number of lines of input text differs greatly from the preset text of the template, the program generates poor results.

The essential reason is that we require the rendered text boundaries to be consistent with the preset text in the template. If the number of lines is too small, the program will use larger line spacing and the text will be sparse; if there are too many lines, the program will use smaller line spacing or font size, making the text difficult to read, affecting the reading experience.

In the future, we can continue to improve the text layout rules of the program, appropriately adjust the optimization goal of text rendering, pay attention to the text rendering effect in various boundary conditions, and make the text part of the generated result more robust.

2) **The template does not support adding multiple sets of pictures, and does not support text other than product title or description.**

   The current template format is determined after researching common posters in the market. Supporting multiple images or texts may make the parameters more complicated, and it will also make the creation and modification of the template more difficult. However, the program allows for further extensions to support rendering of a richer variety of elements, making the resulting posters more diverse.

3) **The relationship between style tags and templates requires manual annotation and cannot be learned automatically.**

   In the existing workflow, after the designer prepares the template, several keywords need to be specified. The program will precisely match the input style tags with the keywords of each template to determine the final template used. However, this process limits the scope of use of templates, because the program cannot yet associate keywords, so that style tags may not necessarily match the appropriate template. The keywords of the template can then be associated to make the matching process smarter.

5.7 Working prototype

We finally tried to build a working prototype to present the results of the whole combinatorial framework, aiming to let the user start with a natural language requirement to get a sketch of the poster idea. As shown in Figure 10, We tentatively name this prototype as AIGD. To start with, the user enters their natural language requirements in the text box on the home page. For this step, we tend to provide some examples of previous user-generated prompts, such as ‘I want a summer coloured, clean-style promotional poster for soda water.’ Once the ‘Generate’ button is clicked, the text is sent to our pre-set combinatorial frame, and the poster sketch is automatically generated. Figure x gives an overview of the prototype's poster sketch page architecture. In addition to the automated generation process, we also give the user the freedom to be inspired by a wider variety of ideas. After selecting the generated poster, the user can modify it via the editing panel. The editing panel consists of five main modules: Generate Results(a), Change Text(b), Change font(c), Change Colours(d) and Change Main Image(e). The Generate Results module generates the resulting visual sketch image. The Change Text module allows you to change the fonts in the poster according to the fonts of our pre-defined style template and to change the relevant product titles and product descriptions. The Change Main Visuals module allows the user to change the main visual image in the sketch, while the colour scheme will be recalculated based on the changes to the main visual image. The Change Colours module is used to change the colour scheme of the sketch based on our calculated colour scheme, e.g. changing similar colours or contrasting colours.

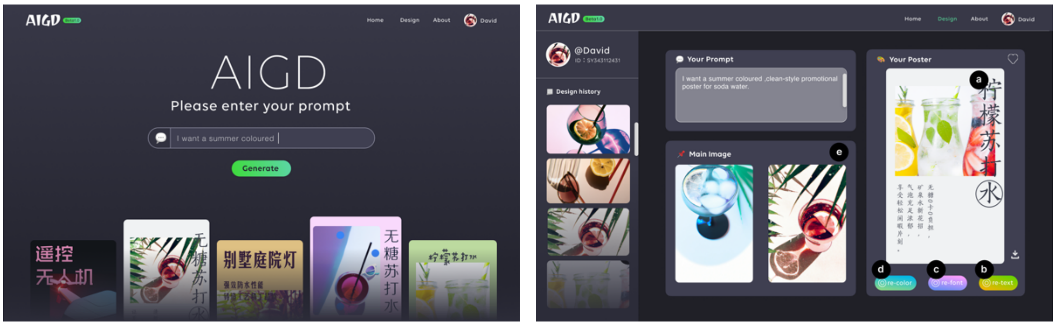


Figure 10: Schematic diagram of the working prototype, the left image is the home page, and the right image is the generated results page.

6 RESEARCH VALIDATION AND EVALUATION

In order to verify the product feasibility and user experience of the prototype, we convened 12 subjects to participate in the experimental evaluation of the generated results.

6.1 Experiment background and conditions

The 12 subjects included 6 professional graphic designers and 6 ordinary users with design service experience. According to their background characteristics, they were divided into two groups of professional graphic designers, namely two groups of design groups: designer group-01 (D1), designer group-02 (D2), in which group D1 is a first-line graphic designer who has worked for about 2 years, and group D2 is a design leader and founder of a design company with many years of work experience.

Two groups of general user groups: general user group-03 (U3), general user group-04 (U4), of which the U3 group is the user group who has purchased graphic design services, but does not have any design skills, and the U4 group are non-professional designers who have tried to create their own designs, with the basic knowledges of design software.

6.2 The specific process of the experiment

1) Experimental process 01:

We asked the 4 groups of subjects to claim the limited design requirements within a limited time: 1h: "The new advertising poster of XX brand soda water, regular vertical size, for online and offline delivery", using AI-aided design tools expand graphic design work.

2) Experiment 02:

After the first stage process, four groups of subjects were asked to create graphic designs in the same limited time (1h) and with the same limited design requirements, using their own traditional design methods.

3) Experiment process 03:

After the two rounds of experiments, the four groups of members were asked to evaluate the overall process of using AI-aided design and production tools in the first round of experiments, specifically for "design efficiency", "design aesthetics", "various solutions". Subjective evaluation of the four dimensions of "sexuality" and "interactive experience", giving each dimension 1-5 points (1 point-very poor, 2 points-poor, 3 points-average, 4 points-good, 5 points- excellent) evaluation.

4) Experimental process 04:

We observed and recorded the different design working methods of the two rounds of experiments, mainly from the time consumed in different design stages and the diversity of output schemes per unit time.

6.3 Analysis of experimental results

As shown in figure 11, from the overall perspective of "design efficiency", "design aesthetics", "program diversity", and "interactive experience", the final weighted average score of all samples is above 3.3 points, " The scores of "design efficiency", "program diversity", and "interactive experience" are all above 4 points, of which the program diversity is the highest at 4.625, indicating that the current prototype's experiential design and generation results have reached an acceptable level for users. Especially in the design efficiency and the output quantity of the design scheme, it has won high user recognition. Therefore, this tool effectively solves the pain points of users for inspiration and production efficiency per unit time as inspiration.

Usability and user experience design evaluation of Demo

Subjective ratings of the subjects

From 1 to 5, the level of satisfaction is from low to high (1-bad, 2-poor, 3-fairl, 4-good, 5-excellent)

Subjects type	Design efficiency improvement	Aesthetics of design Results	Diversity of design schemes	Interactive experience
Designer team 01				
Huang Jin	4	3	4	5
Li Chen	4	2	4	5
Simon	4	3	5	4
TEAM AVERAGE	4	2.6	4.3	4.6
Designer team 02				
Xu Longxiao	3	3	5	4
Olivia	3	2	4	4
Jiang Han	4	4	5	4
TEAM AVERAGE	3.6	3	4.6	4
Design TEAM AVERAGE	3.8	2.8	4.45	4.3
Normal User team 03				
Liu Linjun	5	4	5	3
kim	5	4	5	4
Takashi Yamanaka	5	5	5	5
TEAM AVERAGE	5	4.3	5	4
Normal User team 04				
Chen Hong	5	4	5	3
Wang Yao	4	3	5	5
Li Ying	5	4	4	3
TEAM AVERAGE	4.8	4	4.6	3.6
Normal User TEAM AVERAGE	4.5	4.15	4.8	3.8
Population mean	4.3	3.7	4.625	4.05

Figure 11: For Experiment 1 User Experience Evaluation Visualization Scale

1) The assessment data for "design efficiency" shows an average score of 4.3, which is in the higher satisfaction category overall. The average score for the general user group (U3, U4) in this dimension is 4.8, which is significantly higher than the average score of 3.8 for the designer group (D1, D2). This comparison of scores indicates that the average user without design expertise has a more significant increase in design efficiency.

2) The "design result" assessment is the lowest overall score among the four dimensions, at 3.7, which shows that the current design generation result is only a passing grade compared to

mature and complete solutions on the market, and there is a lot of room for optimisation. However, the prototype is essentially an early inspiration tool for design creators, leading most users to believe that the resulting output is the final result. The data details show that the designer group and the designer groups D1 and D2 scored 2.8. The general user group scored 4.15, which shows that professional designers have more stringent criteria for judging design results. In contrast, general users have a strong sense of freshness due to their output designs and therefore have a greater tolerance for the results.

3) The rating for "design diversity" is 4.625, the highest of all evaluation dimensions. It is worth noting that the average score of all four groups of users is above 4, with the U3 group scoring an average of 5 out of 5. We also note that there is a positive correlation between the scores of "design diversity" and "design efficiency", and we can infer that users believe that "design diversity" promotes "design efficiency". We note a positive correlation between the ratings of 'design diversity' and 'design efficiency', and we can infer that users believe that 'design diversity' has contributed to 'design efficiency'.

4) The overall average score of 4.05 for the interaction experience is good, and the data shows that the average score of professional designers (D1, D2) is 4.3, while the average score of ordinary users (U3, U4) is 3.8. Therefore, we can infer that we should further optimise the interface and interaction process of the product for general users who have not been exposed to design tools or software.

5) Figure 12 presents a comparative analysis of the design output per unit of time of the subjects in Experiment 02 using the traditional and AI-assisted design methods. From the overall results of the four groups, the per capita output of the AI-assisted design method can deliver 2.8 poster designs. In contrast, the result of the traditional design process is 1.07 posters, a three times difference in efficiency. Therefore, we can conclude that the AI-

assisted design method can significantly improve the efficiency of graphic design.

Comparison and analysis of design output per unit time between traditional design model and AI-aided design model

Limited time (1H)

We asked the subjects to design and create with traditional design mode and AI-aided design tools respectively within a limited time (1H) and limited design requirements (soda water advertising poster), and recorded the number and diversity of solutions produced in unit time under the two modes

Subjects type	Ai-aided design process (demo)	Traditional design process
	Number of completed designs	Number of completed designs
Designer team 01		
Huang Jin	4	0.5
Li Chen	3	0.6
Simon	3	0.8
TEAM AVERAGE	**3.3**	**0.63**
Designer team 02		
Xu Longxiao	3	0.8
Olivia	3	1
Jiang Han	5	2
TEAM AVERAGE	3.6	**1.26**
Design TEAM AVERAGE	**3.45**	**0.94**
Normal User team 03		
Liu Linjun	2	/
kim	3	/
Takashi Yamanaka	2	/
TEAM AVERAGE	2.3	/
Normal User team 04		
Chen Hong	2	0.2
Wang Yao	2	0.4
Li Ying	2	0.2
TEAM AVERAGE	2	0.26
Normal User TEAM AVERAGE	**2.15**	**0.13**
Population mean	**2.8**	**1.07**

Figure 12: Comparison of programme outputs for different design approaches in Experiment 2

We continued our analysis from the data details. The group of professional designers (D1, D2) using AI-assisted design tools produced an average of 3.45 poster designs within one hour. At the same time, they only completed one design (0.94) using traditional design methods, a difference in the efficiency of 3.6 times.

The data details also show that the information designers in the D2 group are more familiar with the traditional design process than the D1 junior designers. The difference in efficiency is double, while the number of designs produced by the D1 and D2 groups put into the AI-assisted design process is very similar, 3.3 for D1 and 3.6 for D2. Therefore, we can analyse that AI-aided design tools break the barrier of design efficiency brought about by traditional tool skills. People with creative design ideas are no longer limited to design tools and software limitations and can also produce design solutions quickly.

On the other hand, the general user group lacks professional graphic design work experience. The number of sheets per capita using the traditional design method is 0.13. Members of the U3 group have only purchased design services and do not have any graphic design experience, so the number of outputs per capita is 0. The average output of members of the U4 group is only 0.26 sheets. However, if the data shows that the average output of U3 and U4 under the AI-assisted design method is 2.15 sheets, the difference in efficiency is 8.2 times greater. Thus, we can analyse that compared to the group of professional designers, the average user without professional design experience significantly improved design efficiency.

6.4 Experiment Conclusion

1) The prototype has the core function of good user experience through the overall validation evaluation and analysis. The evaluation results show that generating a variety of design intentions can contribute to the early inspiration of users and increase the productivity of design creators per unit of time.

2) We need to further optimise more details, for example, the natural language description form of user input, the initial design generation effect and quality, and the interface interaction experience for the general user group unfamiliar with the design tool.

7 DISCUSSION

In the following, we discuss several limitations of our system that are identifed from the experiments and user feedback.

1) **Sample size for user research:** The survey data of 117 designers, 17 design companies, and 231 ordinary users collected in the early user research are mainly focused in mainland China. There are only a small number of 12 designer users from the United Kingdom, Italy

and Thailand in the sample, and there is a lack of variety of regions. This may lead to the analysis of user pain points produced by user research, and the research and development opportunities focus on local users in a certain region, and there could be better methods for the research and development of globalization tools under the ideal of this research.

2) **Deviations in Chinese-English translation:** For the sake of equality of results, all the examples of Chinese input and output in this thesis were translated into English using the Youdao Dictionary API translation, which provides a degree of interpretation of the intended translation of the generated results. However, in actual tests, there are cases where the translation is inaccurate and does not express the original meaning in Chinese. This may lead to semantic bias in the multimodal input information of downstream tasks, etc.

3) **Data limitations:** Firstly, because of the limitations of textual data categories, we need to expand to more industries and scenarios. Secondly, due to the uninterpretability of complex machine learning, the model's output will still have a "black box" problem. Thirdly, the model's output is currently mainly in PNG format, and in the future, we need to explore multiple image formats, especially vector graphics. Fourthly, the PSD template of the portfolio generation engine needs to be pre-processed, and the current method still requires manual effort, with long processing times and a lack of supporting design elements, leading to problems such as repetition rates, We can solve this problem in combination with automatic layout generation. We can also build a universal language that translates into something that both human designers and machines can interpret, called a 'design meta-language, and We can build a basic design dataset and a framework for a design Knowledge Graph based on the 'design meta-language'.

4) **Design copyright problem:** Firstly, to address the possible copyright problem in the design content produced by deep algorithms, we should establish an algorithm-based visual similarity detection mechanism to carry out similarity matching detection and determination of preliminary risk of copyright infringement for each user-selected design. Secondly, we need to integrate blockchain encryption technology to achieve direct chain-up of machine design generation results, so as to fundamentally deal with the problem of possible copyright infringement.

5) **Privacy issues:** In order to attain results that better match users' needs, user data may be obtained by machines in the process of human-machine interaction. We should use more secure and reliable encryption methods to avoid leakage of user privacy data. Besides, we should adhere to the principle of technology for benevolence, and exercise a commitment not to use users' data for commercial gain.

6) **Economic issues:** The emergence of increasingly large generated models may eventually have a significant economic influence. Because these models may reduce or replace the work of some people, such as photographers, text editors, illustrators, etc. Guiding machines to

learn may be one of the core jobs of designers in the future. Design is not always rational and logical. Instead, it is more emotional and creative. Design is self-made; through different paradigms and approaches, design always tries to redefine rules and reformulate values and goals.

7) **Research validation and evaluation issues:** The current evaluation method mainly uses the subject's subjective experience evaluation method plus a small amount of objective data record evaluation. The former relies heavily on the experimental designer's overall control of the evaluation dimension, which can easily lead to unreasonable settings for evaluation problems. The small sample size and unrepresentative sample selection resulted in inaccurate evaluation results. At the same time, due to the limitation of physical space and other problems, the background conditions of some subjects are too similar, which may easily lead to the convergence of evaluation results.

Understanding the needs of users is often a costly exercise. AI tools for innovation enable one to actually reduce the attempts to delve into the detailed needs of the user and instead give the user solutions and products that are relevant to their needs along with the appropriate tools. User-centered AI tools for innovation are tailored to specific needs and services. Within these general constraints, users can get the ability to integrate resources, coupled with their own imagination and creativity, and subsequently achieve true freedom of innovation. It is also hoped that research scholars, designers and ordinary users can contribute their own essential and reusable understanding of design, helping to iterate and improve innovative AI tools. This will ultimately lead to a true two-way empowerment of AI and design.

8 CONCLUSION

This study explores the technical path and product solutions of combined intelligent graphic generation design with User-Centered Design Approach. We first explored the different stakeholders during graphic design process through in-depth studies of the current production mechanism of the graphic design industry and summarise the user profiles for the main research objects, that is, the designer group and the general user group with design needs. Then we used qualitative and quantitative research tools and methods to further define the user behaviours. The various behaviours of the two groups of people in the creative process of graphic design are exhaustively listed and clustered to extract high-frequency design behaviours and build a complete blueprint of the design work journey and graphic design services. At last, the core pain points are summarised from the four dimensions of "design tools", "design methods", "design efficiency", and "design experience". The focus of AI-assisted graphic design research is proposed to stimulate design inspiration, lower the threshold of design operation, improve design production efficiency, and allow design creators to spend more time on creativity itself rather than tool operation through AI.

Following our research focus, we first built an innovative generative framework that implements a modular pipeline - models in the pipeline can be replaced by others of same category - for generating poster idea sketches from unstructured inputs based on large-scale training models and regular algorithms. We converted user needs into languages understoodable by machines. Through fully customizable modules, we realized the generation of product descriptions based on user needs, and then obtain a variety of design materials by key word association, and finally combines them with graphic design aesthetic rules. Next, we built a working prototype to show the user the generation process and effects. Then, we evaluate the feasibility of the solution through experimental evaluation of the results of participant participation in prototype generation. The evaluation results show that the prototype can help stimulate the early creative inspiration of users by producing a variety of graphic design intentions, and improve the design production efficiency of design creators per unit time, especially for ordinary users who do not have professional graphic design capabilities. However, the experimental data also shows that there is still room for further optimization in the style effect and quality of the initial design generation and the interactive experience of the prototype product.

To sum up, through user research, technology integration, prototype construction, verification testing, and evaluation, we have completed the initial expected research goals. Through the integration of artificial intelligence and graphic design, a feasible process of AI-assisted graphic design is constructed to help stimulate the creative inspiration of design and entrepreneurial practitioners and improve design efficiency. We hope that it can help us break the monopoly of design skills brought by professional design tools, so that ordinary people without any design skills can also obtain their own graphic design ideas.

ACKNOWLEDGEMENTS

It would not have been possible to accomplish this thesis without the support and help of my supervisor and friends. First of all, I would like to thank my supervisor Dr Louis McCallum who encouraged and supported my work throughout my MSc study. Many thanks for giving me constructive advice, consideration and a joyful environment to explore new ideas. His continuous enthusiasm, patient guidance, and promotion have made my precious study time easier and happier.

I would like to thank all my friends and classmates during the past journey. Many thanks to my boyfriend, Shengyu Huang, for his love and his suggestions on my thesis.

Thank you to all those who participated in my research process.

ual: creative computing institute

REFERENCE

aletts (2021). Disco Diffusion. [online] GitHub. Available at: https://github.com/alembics/disco-diffusion [Accessed 30 Jul. 2022].

Arroyo, D., Postels, J. and Tombari, F. (2021). Variational Transformer Networks for Layout Generation. [online] Available at: https://arxiv.org/pdf/2104.02416v1.pdf [Accessed 7 Oct. 2022].

Bharadhwaj, H., Park, H. and Lim, B.Y. (2018). RecGAN. Proceedings of the 12th ACM Conference on Recommender Systems. doi:10.1145/3240323.3240383.

CairX (2020). extcolors: Extract colors from an image. Colors are grouped based on visual similarities using the CIE76 formula. [online] PyPI. Available at: https://pypi.org/project/extcolors/ [Accessed 8 Oct. 2022].

Chen, L., Wang, P., Shi, F., Han, J. and Childs, P. (2018). A COMPUTATIONAL APPROACH FOR COMBINATIONAL CREATIVITY IN DESIGN. pp.1815–1824. doi:10.21278/idc.2018.0375.

Cheng, Z., Yang, Q. and Sheng, B. (2016). Deep Colorization. arXiv:1605.00075 [cs]. [online] Available at: https://arxiv.org/abs/1605.00075.

Computer Hope (2019). What is a GUI (Graphical User Interface)? [online] Computerhope.com. Available at: https://www.computerhope.com/jargon/g/gui.htm.

Crowson, K., Biderman, S., Kornis, D., Stander, D., Hallahan, E., Castricato, L. and Raff, E. (2022). VQGAN-CLIP: Open Domain Image Generation and Editing with Natural Language Guidance. arXiv:2204.08583 [cs]. [online] Available at: https://arxiv.org/abs/2204.08583.

Dou, Q., Zheng, X.S., Sun, T. and Heng, P.-A. (2019). Webthetics: Quantifying webpage aesthetics with deep learning. International Journal of Human-Computer Studies, 124, pp.56–66. doi:10.1016/j.ijhcs.2018.11.006.

dreams, S. (n.d.). Google Colaboratory. [online] colab.research.google.com. Available at: https://colab.research.google.com/github/alembics/disco-diffusion/blob/main/Disco_Diffusion.ipynb [Accessed 26 Jun. 2022].

Eric von Hippel (2001). PERSPECTIVE: User toolkits for innovation. Journal of Product Innovation Management, [online] 18, pp.247–257. doi:https://doi.org/10.1016/S0737-6782(01)00090-X.

Gatys, L.A., Ecker, A.S. and Bethge, M. (2015). A Neural Algorithm of Artistic Style. [online] arXiv.org. Available at: https://arxiv.org/abs/1508.06576.

Guo, S., Jin, Z., Sun, F., Li, J., Li, Z., Shi, Y. and Cao, N. (2021). Vinci: An Intelligent Graphic Design System for Generating Advertising Posters. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. doi:10.1145/3411764.3445117.

Gupta, K., Lazarow, J., Achille, A., Davis, L., Mahadevan, V. and Shrivastava, A. (2021). LayoutTransformer: Layout Generation and Completion with Self-attention. arXiv:2006.14615 [cs]. [online] Available at: https://arxiv.org/abs/2006.14615 [Accessed 7 Oct. 2022].

Han, J., Shi, F., Chen, L. and Childs, P. (2018). The Combinator – a computer-based tool for creative idea generation based on a simulation approach. Design Science, 4. doi:10.1017/dsj.2018.7.

Huang, F. and Canny, J.F. (2019). Sketchforme: Composing Sketched Scenes from Text Descriptions for Interactive Applications. arXiv:1904.04399 [cs]. [online] Available at: https://arxiv.org/abs/1904.04399 [Accessed 6 Oct. 2022].

huggingface.co. (n.d.). IDEA-CCNL/Randeng-BART-139M-SUMMARY · Hugging Face. [online] Available at: https://huggingface.co/IDEA-CCNL/Randeng-BART-139M-SUMMARY [Accessed 8 Oct. 2022].

huggingface.co. (n.d.). IDEA-CCNL/Randeng-Pegasus-238M-Summary-Chinese · Hugging Face. [online] Available at: https://huggingface.co/IDEA-CCNL/Randeng-Pegasus-238M-Summary-Chinese [Accessed 8 Oct. 2022].

Jahanian, A., Liu, J., Lin, Q., Tretter, D., Eamonn O'Brien-Strain, Seungyon Claire Lee, Lyons, N. and Allebach, J.P. (2013). Automatic design of colors for magazine covers.

Jung, S.-G., Salminen, J., Kwak, H., An, J., Jansen, B. and Kwak, J. (2018). Automatic Persona Generation (APG): A Rationale and Demonstration CCS CONCEPTS • Human-centered computing ⟶ Human computer interaction (HCI) ⟶ Empirical studies in HCI KEYWORDS Personas; User Experience Research; User Analytics ACM Reference format. [online] Available at:

ual: creative computing institute

https://secureservercdn.net/160.153.138.243/33e.495.myftpupload.com/wp-content/uploads/2018/08/automatic-persona-generation-a-rationale.pdf?time=1665001337 [Accessed 6 Oct. 2022].

Kim, J., Maher, M. and Siddiqui, S. (2021). Collaborative Ideation Partner: Design Ideation in Human-AI Co-creativity. Proceedings of the 5th International Conference on Computer-Human Interaction Research and Applications. [online] doi:10.5220/0010640800003060.

Kong, S., Shen, X., Lin, Z., Mech, R. and Fowlkes, C. (2016). Photo Aesthetics Ranking Network with Attributes and Content Adaptation. [online] Available at: https://arxiv.org/pdf/1606.01621.pdf [Accessed 7 Oct. 2022].

Kong, X., Jiang, L., Chang, H., Zhang, H., Hao, Y., Gong, H. and Essa, I. (2022). BLT: Bidirectional Layout Transformer for Controllable Layout Generation. arXiv:2112.05112 [cs]. [online] Available at: https://arxiv.org/abs/2112.05112 [Accessed 29 Sep. 2022].

Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), pp.84–90. doi:10.1145/3065386.

Lee, H.-Y., Jiang, L., Essa, I., Le, P., Gong, H., Yang, M.-H. and Yang, W. (2020). Neural Design Network: Graphic Layout Generation with Constraints. [online] Available at: https://arxiv.org/pdf/1912.09421.pdf [Accessed 7 Oct. 2022].

Li, J., Yang, J., Hertzmann, A., Zhang, J. and Xu, T. (2021). LayoutGAN: Synthesizing Graphic Layouts With Vector-Wireframe Adversarial Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(7), pp.2388–2399. doi:10.1109/tpami.2019.2963663.

logCong (2021). GPT2-NewsTitle. [online] GitHub. Available at: https://github.com/liucongg/GPT2-NewsTitle.

Lu, X., Lin, Z., Shen, X., Mech, R. and Wang, J.Z. (2015). Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. pp.990–998. doi:10.1109/ICCV.2015.119.

McKinsey & Company. (2019). McKinsey Global Institute. [online] Available at: https://www.mckinsey.com/mgi/overview.

Mihalcea, R. and Tarau, P. (2004). TextRank: Bringing Order into Text. [online] ACLWeb. Available at: https://aclanthology.org/W04-3252/.

NightCafe Creator. (n.d.). AI Generated Art Prints. [online] Available at: https://nightcafe.studio/.

Norman, D.A. and Draper, S.W. (1986). User centered system design : new perspectives on human-computer interaction. Hillsdale, N.J.: Lawrence Erlbaum.

Perera, D. and Zimmermann, R. (2019). CnGAN: Generative Adversarial Networks for Cross-network user preference generation for non-overlapped users. The World Wide Web Conference on - WWW '19, [online] pp.3144–3150. doi:10.1145/3308558.3313733.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C. and Openai, M. (2022a). Hierarchical Text-Conditional Image Generation with CLIP Latents. [online] Available at: https://arxiv.org/pdf/2204.06125.pdf.

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C. and Openai, M. (2022b). Hierarchical Text-Conditional Image Generation with CLIP Latents. [online] Available at: https://arxiv.org/pdf/2204.06125.pdf.

Sheng, K., Dong, W., Ma, C., Mei, X., Huang, F. and Hu, B.-G. (2018). Attention-based Multi-Patch Aggregation for Image Aesthetic Assessment. Proceedings of the 26th ACM international conference on Multimedia. [online] doi:10.1145/3240508.3240554.

Singh, D., Rajcic, N., Colton, S. and McCormack, J. (2019). Camera obscurer: Generative art for design inspiration.

STEVENKEJJAD (2020). subreddits. [online] www.reddit.com. Available at: https://www.reddit.com/r/ [Accessed 6 Oct. 2022].

Wang, J., Yu, L., Zhang, W., Gong, Y., Xu, Y., Wang, B., Zhang, P. and Zhang, D. (2017). IRGAN: A Minimax Game for Unifying Generative and Discriminative Information Retrieval Models. Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, [online] pp.515–524. doi:10.1145/3077136.3080786.

Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X. and He, X. (2017). AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks. [online] arXiv.org. Available at: https://arxiv.org/abs/1711.10485.

Xu, Z.-Q.J., Zhang, Y., Luo, T., Xiao, Y. and Ma, Z. (2020). Frequency Principle: Fourier Analysis Sheds Light on Deep Neural Networks. Communications in Computational Physics, [online] 28(5), pp.1746–1767. doi:10.4208/cicp.OA-2020-0085.

Yang, X., Mei, T., Xu, Y.-Q., Rui, Y. and Li, S. (2016a). Automatic Generation of Visual-Textual Presentation Layout. ACM Transactions on Multimedia Computing, Communications, and Applications, 12(2), pp.1–22. doi:10.1145/2818709.

Yang, X., Mei, T., Xu, Y.-Q., Rui, Y. and Li, S. (2016b). Automatic Generation of Visual-Textual Presentation Layout. ACM Transactions on Multimedia Computing, Communications, and Applications, 12(2), pp.1–22. doi:10.1145/2818709.

You, W.-T., Sun, L.-Y., Yang, Z.-Y. and Yang, C.-Y. (2019). Automatic advertising image color design incorporating a visual color analyzer. Journal of Computer Languages, [online] 55, p.100910. doi:10.1016/j.cola.2019.100910.

Zen, M. and Vanderdonckt, J. (2014). Towards an evaluation of graphical user interfaces aesthetics based on metrics. 2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS). doi:10.1109/rcis.2014.6861050.

Zhang, J., Yu, J., Zhang, K., Zheng, X.S. and Zhang, J. (2017). Computational Aesthetic Evaluation of Logos. ACM Transactions on Applied Perception, 14(3), pp.1–21. doi:10.1145/3058982.

Zhao, J., Han, J., Shao, L. and Snoek, C.G.M. (2019). Pixelated Semantic Colorization. arXiv:1901.10889 [cs]. [online] Available at: https://arxiv.org/abs/1901.10889 [Accessed 6 Oct. 2022].

周子洪周志斌 (2020). 人工智能赋能数字创意设计:进展与趋势. 计算机集成制造系统, [online] 26(10), p.2603. doi:10.13196/j.cims.2020.10.001.