Transactions on Computational and Scientific Methods | Vo. 5, No. 3, 2025
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Generative UI Design with Diffusion Models: Exploring Automated Interface Creation and Human-Computer Interaction

Shixiao Wang[1], Runsheng Zhang[2], Xue Shi[3]
[1]School of Visual Arts, New York, USA
[2]University of Southern California, Los Angeles, USA
[3]New York University, New York, USA
*Corresponding author: Xue Shi, xuexues9998@gmail.com

Abstract: This study focuses on generating UI interfaces based on the diffusion model, aiming to improve the human-computer interaction experience through generative models. As an emerging deep-learning technology, the diffusion model can effectively generate high-quality images by simulating the gradual introduction and removal of noise. Traditional UI design usually relies on manual design and pre-defined templates, while automated design based on the diffusion model can generate creative, structured, and diverse UI interfaces. This study uses the diffusion model to model the UI interface and generates interface elements that meet the design requirements by guiding the model from noise to clear images. The experimental results show that the diffusion model performs well in generating patterns and background parts in UI design and can successfully simulate complex layouts and visual effects. However, despite the satisfactory effect of pattern generation, the model still faces certain challenges when generating text parts in the UI interface. The text content sometimes appears unclear, blurred or garbled, which affects the readability and overall effect of the generated interface. To address this problem, future research can explore how to further optimize the model and improve the accuracy and clarity of text generation. Overall, the UI interface generation technology based on the diffusion model shows great potential, which can provide designers with new creative tools and promote the development of the UI design field. With the improvement of algorithms and computing power, the diffusion model is expected to become an important method for automated UI design, promoting further optimization and innovation of the human-computer interaction experience.

Keywords: Diffusion model, UI interface generation, human-computer interaction, automated design

1. Introduction

With the rapid advancement of information technology and artificial intelligence, user interface (UI) design has become increasingly vital in modern software development. Traditional UI design relies heavily on manual methods and templated approaches [1]. While these methods can address functional requirements to some extent, they often fall short in aesthetics, interactivity, and personalization. This is particularly evident in complex application scenarios, where designers must handle numerous user demands and interface optimization tasks [2]. As a result, traditional design methods are struggling to keep pace with the rapidly evolving market needs. Consequently, finding innovative technical solutions to improve UI design efficiency and enhance user experience has become a critical research topic in the field of human-computer interaction.

Diffusion models, a type of deep generative model, have emerged in recent years with powerful image-generation capabilities. By simulating the gradual diffusion of data distributions, these models can capture intricate image details and structural features when generating new images. Unlike traditional image generation methods, diffusion models not only account for global structures but also preserve local details effectively. This makes them particularly advantageous in image generation tasks. Recently, diffusion models have been applied in various fields, including image restoration, style transfer, and image generation [3]. However, their application in UI design remains in the exploratory phase [4].

Research on generating UI interfaces using diffusion models offers a breakthrough beyond traditional UI design methods. It also enhances design efficiency through automated algorithmic generation. By learning from large datasets of UI interfaces, diffusion models can identify underlying patterns in design styles, layouts, and color combinations. This provides designers with creative inspiration and facilitates rapid adaptation to changing user needs. As artificial intelligence technology continues to advance, machine learning algorithms for UI design have become a growing trend. This is particularly evident in scenarios that require personalization and efficient iteration, where diffusion models have demonstrated significant potential [5].

Moreover, as human-computer interaction experiences continue to improve, user expectations for interface aesthetics, interactivity, and responsiveness are rising. A UI is no longer just a medium for displaying information; it is an essential bridge for user-system interaction. UI interfaces generated by diffusion models can more accurately capture user needs and aesthetic trends, thereby enhancing the overall user experience [6]. Particularly with the widespread use of devices such as smartphones and smart home systems, the innovation and optimization of UI interfaces have become even more crucial. The challenge of how to enhance the intelligence and personalization of UI design through algorithms has become a key topic in human-computer interaction research [7].

Researching algorithms for generating UI interfaces based on diffusion models offers a valuable complement to existing design methods. It also provides a key avenue for advancing human-computer interaction technology. By innovating algorithms, automated UI design can be realized, reducing manual intervention, enhancing design efficiency, and lowering development costs. This approach supports the future intelligent and personalized evolution of UI design.

2. Related work

2.1 Diffusion Model

Diffusion models have made significant progress in the field of generative models in recent years and have demonstrated their strong potential in multiple visual tasks. Initially, diffusion models were used for image-generation tasks, and researchers found that by gradually introducing noise and then denoising it in reverse, high-quality images can be effectively generated. Compared with traditional generative adversarial networks (GANs), diffusion models can avoid problems such as mode collapse during training, so they have achieved good results in the field of image generation. The core idea of this model is to gradually diffuse noise and eventually restore high-quality images, which has attracted widespread attention for its performance in generation tasks.

In addition to image generation, diffusion models have also been applied in other fields, especially in image-to-image conversion tasks. Researchers have tried to combine diffusion models with conditional generation models to further enhance their capabilities in image reconstruction and image restoration. This combination enables diffusion models to generate output images with high visual quality under different input conditions, such as super-resolution tasks from low resolution to high resolution or denoising and image restoration tasks. These advances provide a theoretical basis for the application of diffusion models in UI interface generation and other design fields.

The advantages of diffusion models extend beyond image generation. Recent studies have explored their potential in text generation for natural language processing (NLP) tasks, as well as in multimodal generation, where images and text are combined. Although these efforts are still in the early stages, they highlight the promising capabilities of diffusion models in generating complex structures and content. In the future, they may offer enhanced generation capabilities for tasks like UI design.

In the field of UI design, the study of diffusion models is relatively new, but its application prospects are worth looking forward to. The current UI design relies more on human designers and preset templates. Diffusion models provide new ideas for automated UI design through their generation capabilities. Although existing diffusion models can achieve good results in generating image elements, there is still room for improvement in text generation and overall interface layout. Future research may focus on improving the accuracy and clarity of the text part of the model-generated UI interface and its combination with graphic elements, thereby promoting the further development of UI interface generation technology.

2.2 UI interface generation

In recent years, deep learning-based methods for UI design generation have gradually become a research hotspot in the field of human-computer interaction. Early studies primarily focused on using Generative Adversarial Networks (GANs) to generate UI layout designs. Some methods achieved this by designing specific network architectures to generate UI component layouts that met certain standards. By learning from a large number of UI design samples, these methods were able to generate reasonably visually coherent interfaces to some extent. However, GANs face challenges such as mode collapse and a lack of diversity during the training process, which remain obstacles in UI design generation [8].

With the significant breakthroughs of diffusion models in the field of image generation, more and more research has shifted toward applying diffusion models to UI design generation [9]. Unlike GANs, diffusion models generate images gradually through a reverse diffusion process. This allows them to better preserve details and structural integrity, avoiding some of the limitations found in traditional methods. Some researchers have attempted to apply diffusion models to the automatic generation of UI interface elements. Studies show that diffusion models achieve higher generation quality when producing complex, layered interface designs, and they are capable of effectively simulating design styles and user preferences [10-11].

Although research on UI interface generation based on diffusion models has made some progress, current work primarily focuses on generating interface components. Little has been done to explore how to integrate UI with business logic effectively using this technology. Traditional UI design methods typically separate the interface from business logic. However, enabling intelligent algorithms to automatically integrate business logic into interface design remains an important research direction. Moreover, improving the computational efficiency of generation algorithms and enabling quick responses to user needs in practical applications are key challenges for the widespread adoption of this technology in commercial products [12].

3. Method

In this study, we proposed an algorithm for generating UI interfaces based on a diffusion model, aiming to improve the automation and efficiency of UI design through deep learning methods. First, we regard the UI interface as an image generation problem, generate UI interface images through a diffusion model, and combine specific constraints to optimize the generation process.

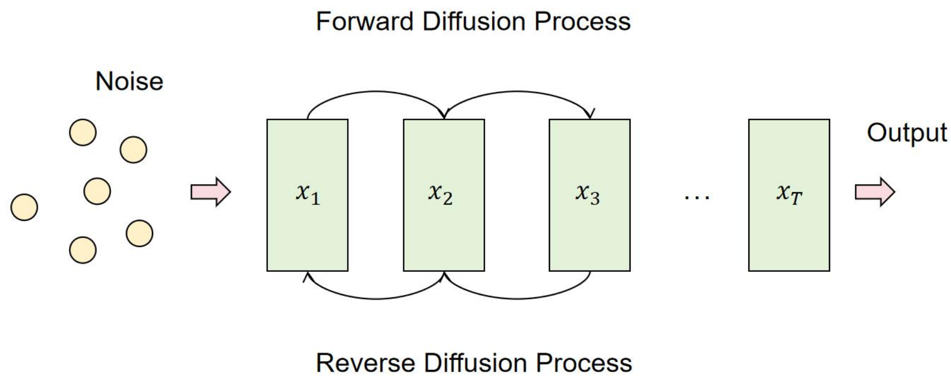

Figure 1. Diffusion model forward and backward example diagram

As shown in Figure 1, the diffusion model first gradually transforms the input noise into a series of states through the forward diffusion process, which gradually carry noise until they finally become completely noisy states. Then, the reverse diffusion process gradually denoises and restores the data through reverse steps, and finally outputs the generated image. The goal of this process is to generate high-quality images or UI interface designs by learning the reverse process from noise to data.

In the specific experimental process, we first defined the target distribution $p_{data}(x)$ of UI design images, where x represents the UI design image and p_{data} is the real distribution of image data. The diffusion process gradually destroys the data distribution by introducing noise, and the reverse process gradually restores the structure and details of the image through the model.

The generation process of the diffusion model usually includes two stages: forward diffusion and reverse denoising[13-15]. Assuming that we have a set of training data x_0 to represent the initial UI design image, we convert the data into noise through the forward diffusion process, which can be expressed as:

$$q(x_t \mid x_{t-1}) = N(x_t; \sqrt{1-\beta_t} x_{t-1}, \beta_t I)$$

Where β_t is the noise adjustment parameter of the diffusion step t, N represents the Gaussian distribution, and x_t is the image after diffusion. The goal of the forward process is to gradually transform the input image from the true distribution to noise through a series of steps. Then, in the reverse denoising process, we learn the denoising process by optimizing the model. The reverse process can be expressed as:

$$p_\theta(x_{t-1} \mid x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I)$$

Where $\mu_\theta(x_t, t)$ is the denoised mean of the model output and σ_t^2 is the noise variance associated with the time step. By maximizing the log-likelihood function $\log p_\theta(x_0)$, we can train the model to learn the ability to generate UI design images.

In order to improve the quality and efficiency of generating UI interfaces, we introduced UI-specific constraints into the diffusion model [16]. Specifically, we added an additional loss function during the training process, which aims to optimize the layout rationality and visual effects of the UI interface. Assuming y represents the layout structure and attributes in the design, we can introduce this constraint into the model through the conditional generation method:

$$p_\theta(x_0 \mid y) = \int p_\theta(x_0 \mid x_T) p_\theta(x_T \mid y) d_{xT}$$

Among them, y is the UI layout parameter generated by user needs or design specifications, and x_T is the intermediate noise variable after multiple diffusions. In this way, we can control the generated UI interface to meet the preset design specifications and user needs.

In addition, in order to improve the stability and controllability of the generation process, we adopted an adaptive noise adjustment method [17]. At each time step t, we dynamically adjust the noise level by calculating the gradient information of the current generated image, making the generation process more stable. The optimization objective function is as follows:

$$L_{total} = E_q[\| x_t - f_\theta(x_{t-1}) \|^2] + \lambda[\| x_t - x_t^{'} \|^2]$$

Among them, $f_\theta(x_{t-1})$ is the predicted output of the model, λ is the hyperparameter of the balanced noise adjustment term, and $x_t^{'}$ is the generated image based on the current noise level. By minimizing this loss function, we can effectively improve the quality and consistency of UI generation.

4. Experiment

4.1 Datasets

The WebUI dataset is a diverse and open resource library specifically designed for the study of human-computer interaction interface design, aiming to provide a rich set of UI design styles and layout patterns. The dataset contains a large number of screenshots of web pages and mobile application interfaces, covering different industries and functions, such as e-commerce, social platforms, financial services, etc. Therefore, the WebUI dataset can not only help researchers explore the diversity of interface design, but also provide design inspiration and ideas for different user needs.

This dataset is particularly suitable for research on UI generation and optimization using machine learning and deep learning methods. Its rich image data can provide enough training samples for the diffusion model to generate diverse interface designs that conform to user interaction habits. The WebUI dataset also includes layouts and elements of different styles. Researchers can filter out relevant parts according to specific research objectives to improve the accuracy and effect of the algorithm.

By using the WebUI dataset, researchers can better understand and simulate the interaction process between users and interfaces. It provides strong support for exploring how to generate UI interfaces that meet human-computer interaction needs through algorithms, making deep learning-based automated design technology possible, and further promoting the optimization and innovation of human-computer interaction experience.

4.2 Experimental Results

Table 1: Experimental results

Evaluation Dimensions	**score**	**illustrate**
Interface aesthetics	4.32	The interface design was rated highly for its visual appeal, with users appreciating the color scheme, layout, and overall design aesthetics. Most users found the interface to be modern and visually engaging, although a few noted that the contrast between some elements could be improved for better visibility.
Ease of use	4.11	Users found the interface relatively easy to navigate, with clear instructions and intuitive controls. However, some reported occasional confusion in locating certain features, which suggests that while the design is user-friendly, further refinement could enhance clarity and reduce the learning curve.
Functional accessibility	4.25	The majority of users had no trouble accessing key features and functions of the interface. The layout was considered effective for facilitating easy navigation to important sections. A few users, however, pointed out that additional shortcuts or more prominent feature placement could further streamline accessibility.
Customer satisfaction	4.00	Overall satisfaction was positive, with most users reporting a favorable experience. The ease of use and aesthetic appeal contributed significantly to user contentment. Nevertheless, some users suggested improvements in the response speed and minor design adjustments to enhance the overall experience.

As can be seen from Table 1, users are generally satisfied with the interface design, especially in terms of interface aesthetics and functional accessibility. Most users think that the interface is modern, the color and layout design are attractive and can provide a good visual experience. However, users' scores for ease of use are slightly lower. Although the interface is generally intuitive, there are still some unclear points in the positioning and operation of specific functions, which may affect the user experience of new users. Improving guidance and reducing operational complexity may further optimize the score of this dimension.

Although customer satisfaction is generally positive, its score (4.00) is slightly lower than other dimensions, reflecting that although users generally think that the interface is easy to operate and beautiful, there is still room for improvement. In particular, response speed and some design details, such as contrast of elements and quick access to functions, are all concerns in user feedback. In order to improve user experience, further optimization may be needed in technical implementation, such as improving the response speed of the system or simplifying complex interactive elements in the interface.

Secondly, we give a graph of the decrease in loss function during the training process, as shown in Figure 2.

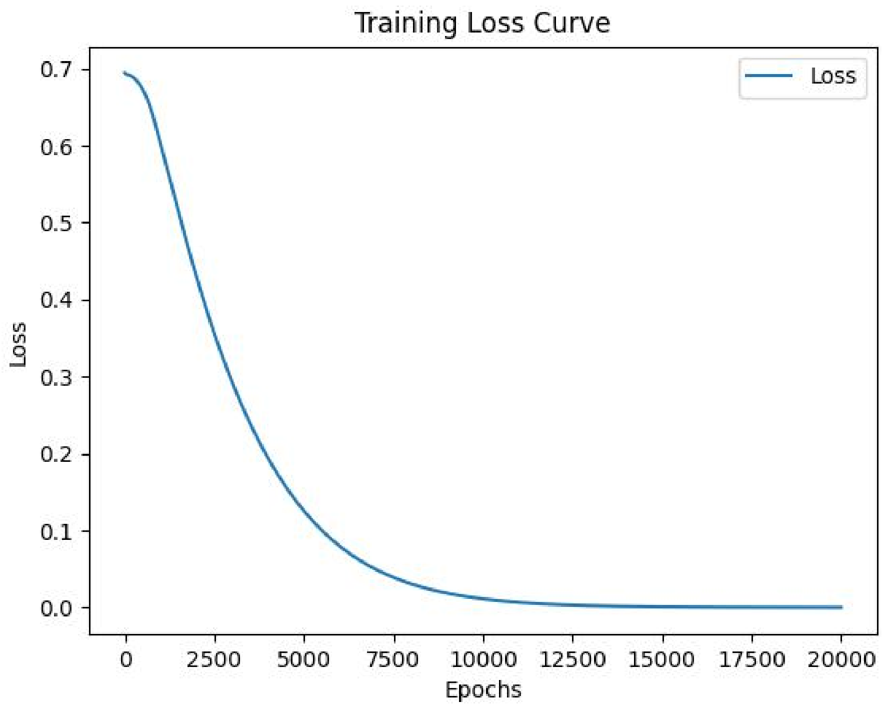

Figure 2. Training loss drop graph

Figure 2 shows the decline curve in the loss function during training. It can be seen that the loss value drops rapidly in the first few epochs, which indicates that the model quickly learns some effective features in the early stage. The rate of decline of the loss function gradually slows down and enters a stable stage, close to about 0.1, indicating that the model gradually converges and tends to the optimal solution in the later stage.

From the figure, it can be inferred that within the first few thousand epochs of model training, the network has a high learning efficiency and can quickly improve its prediction ability. However, as training continues, the loss value tends to stabilize and the rate of further reduction slows down significantly. This phenomenon usually indicates that the model is close to the optimal solution, there is limited room for further optimization, and there may be a risk of overfitting.

The smooth curve in the figure may also imply good stability of the training process, and there is no significant fluctuation or repeated change in the training process, which is usually a sign of a healthy training process. In order to further optimize the model, more parameter adjustments or other technical means, such as regularization or adjusting the learning rate, may be required to help the model converge more effectively.

This image shows the process of removing noise from a UI interface through a diffusion model. The change from top to bottom shows that the noise is gradually removed from the initial interference state, and the interface gradually returns to a clear state. Initially, the noise seriously interferes with the clarity of the image. As the denoising process progresses, the noise gradually disappears, and the details and structure of the UI interface become more and more obvious, until the final image is almost restored to its original noise-free state.

Under the framework of the diffusion model, this denoising process simulates the gradual dissipation of noise, helping the model learn how to restore a clear interface image under different noise intensities. The diffusion model not only retains the basic structure of the UI interface during the denoising process, but also effectively eliminates the interference caused by noise, thereby improving the visual quality of the image and user experience. This process not only shows the removal of noise, but also emphasizes the importance of restoring clarity and details in complex environments.

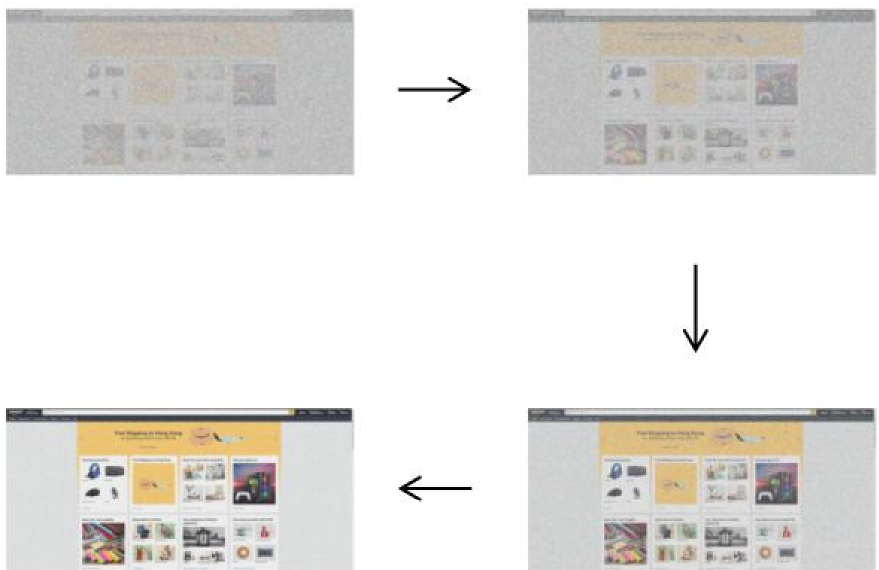

Figure 3. The process of removal of noise to the UI interface

Finally, this paper presents the UI interface diagram generated by the diffusion model, as shown in Figure 4.

As can be seen from the image, the UI interface generated by the diffusion model performs well in pattern design, and the details and visual effects are relatively clear and attractive. The icons and background elements in the image show a certain aesthetic effect, and the colors and design style also appear modern and interesting. This shows that the diffusion model can better capture complex graphic structures and visual elements when generating images, and is suitable for generating intuitive and expressive UI designs.

However, in the generation of text, the model still has certain shortcomings. The text content in the image looks a bit messy and difficult to read, the generated text part is not clear enough, and sometimes garbled or illogical characters appear. This shows that in terms of text generation, the diffusion model needs to be further improved, especially in terms of text structure and typesetting, and the model's understanding of language and symbols needs to be strengthened.

Overall, although the current diffusion model has achieved good results in graphic design, there is still room for improvement in the text performance of the UI interface. In the future, the model's performance in text generation can be further improved by optimizing the model's training data set and adding more text processing capabilities. Through this improvement, the diffusion model is expected to generate more complete UI designs in the future and enhance the overall human-computer interaction experience.

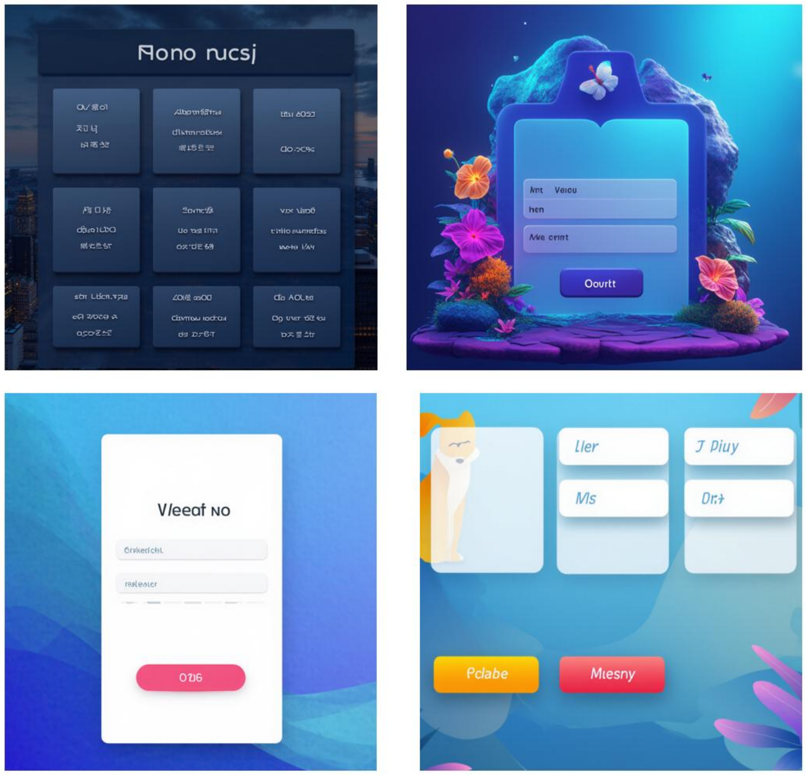

Figure 4. Schematic diagram of the generated UI interface

5. Conclusion

This paper studies the algorithm for generating UI interfaces based on the diffusion model and explores its application in improving the human-computer interaction experience. The experimental results show that the diffusion model can effectively generate UI designs with high visual quality, especially in the generation of patterns and background elements. However, the model still has certain shortcomings in text generation, especially the clarity and structure of the text content, which affects the overall readability and user experience of the UI interface.

With the deepening of research, the future diffusion model can improve its performance in text processing by introducing more text generation optimization technologies. For example, the combination of natural language processing (NLP) technology and image generation models may achieve more accurate text generation and layout arrangements in the future. By further optimizing the training data set, the model can improve the understanding and generation ability of language content while maintaining the aesthetics of the image, avoiding garbled or irregular text forms. In addition, the robustness and adaptability of the model are also the focus of future research. How to maintain the quality of the generated UI interface under different environments and noise conditions and avoid being affected by external interference will be an important direction for improving the human-computer interaction experience. Combined with cutting-edge technologies such as multimodal learning and self-supervised learning, the diffusion model is expected to show higher flexibility and responsiveness in complex application scenarios. In short, the UI interface generation technology based on the diffusion model has shown its great potential, but there are still many aspects that need to be continuously optimized and improved. In the future, with the improvement of computing power and the continuous development of algorithms, the diffusion model is expected to bring more intelligent, accurate and efficient UI design solutions to the field of human-computer interaction, and promote the human-computer interaction experience into a new stage.

References

[1] Sun, Layla, Mengmeng Qin, and Benji Peng. "Llms and diffusion models in ui/ux: Advancing human-computer interaction and design." OSF Preprints (2024).

[2] Chen, Junxing, Yi Zou, and Xu Chu. "DIFFUSUP: A graphical user interface (GUI) software for diffusion modeling." Applied Computing and Geosciences 22 (2024): 100157.

[3] Yuan, Mingyue, Jieshan Chen, and Aaron Quigley. "MAxPrototyper: A Multi-Agent Generation System for Interactive User Interface Prototyping." arXiv preprint arXiv:2405.07131 (2024).

[4] Chávez, Poma, and Wilfredo Ticona. "Implementation of Text-to-Image Generators in the Development of the Usability Interface for the Construction of a Web Page." 2024 14th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2024.

[5] Duan, S. (2024). Systematic Analysis of User Perception for Interface Design Enhancement. Journal of Computer Science and Software Applications, 5(2).

[6] Tian, Yingtao. "DiffCJK: Conditional Diffusion Model for High-Quality and Wide-coverage CJK Character Generation." arXiv preprint arXiv:2404.05212 (2024).

[7] Kobayashi, Masakazu, and Katsuyoshi Kume. "Design Generation Using Stable Diffusion and Questionnaire Survey." (2024).

[8] Feng, Yutong, et al. "Ranni: Taming text-to-image diffusion for accurate instruction following." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

[9] Sun, Q. (2025). Spatial Hierarchical Voice Control for Human-Computer Interaction: Performance and Challenges. Journal of Computer Technology and Software, 4(1).

[10] Kabir, Ahmed Imran, et al. "Empowering Local Image Generation: Harnessing Stable Diffusion for Machine Learning and AI." Informatica Economica 28.1 (2024).

[11] Todi, K., Bailly, G., Leiva, L., & Oulasvirta, A. (2021, May). Adapting user interfaces with model-based reinforcement learning. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-13).

[12] Nguyen, Luan, and Minh Hoang. "Building a fullstack mobile application with flutter and stable diffusion model." (2024).

[13] Liu, F., & Su, J. (2004). An online feature learning algorithm using HCI-based reinforcement learning. In Advances in Neural Networks–ISNN 2004: International Symposium on Neural Networks, Dalian, China, August 2004, Proceedings, Part I 1 (pp. 293-298). Springer Berlin Heidelberg.

[14] Arzate Cruz, C., & Igarashi, T. (2020, July). A survey on interactive reinforcement learning: Design principles and open challenges. In Proceedings of the 2020 ACM designing interactive systems conference (pp. 1195-1209).

[15] Zhang, C., Wang, S., Aarts, H., & Dastani, M. (2021). Using cognitive models to train warm start reinforcement learning agents for human-computer interactions. arXiv preprint arXiv:2103.06160.

[16] Sun, Q., & Duan, S. (2025). User Intent Prediction and Response in Human-Computer Interaction via BiLSTM. Journal of Computer Science and Software Applications, 5(3).

[17] Gaspar-Figueiredo, D., Fernández-Diego, M., Nuredini, R., Abrahão, S., & Insfrán, E. (2024, June). Reinforcement Learning-Based Framework for the Intelligent Adaptation of User Interfaces. In Companion Proceedings of the 16th ACM SIGCHI Symposium on Engineering Interactive Computing Systems (pp. 40-48).