



DESIGN FOR DIVERSE BODY SHAPES WITH AI

Exploring AIGC Applications for DINED Mannequin Generation





Author

Yunzhao Ma | 5242819



TU Delft

DINED

anthropometric
database

Delft University of Technology

Faculty of Industrial Design Engineering (IDE)

MSc. Design for Interaction

Master Graduation Thesis

Design for Diverse Body Shapes with AI

Exploring AIGC Applications for DINED Mannequin Generation

(AIGC: Artificial Intelligence Generated Content)

Supervisory Team

(HCD: Department of Human-Centered Design)

Chair | Dr. Huysmans, T.

Toon Huysmans

IDE - HCD - Applied Ergonomics & Design

Expert in Digitization / The 3D/4D Anthropometry Lead of DINED

Mentor | Dr. Lomas, J.D.

Derek Lomas

IDE - HCD - Design Aesthetics

Expert in Positive AI / Bring Humanist Values Into AI Systems

July 2023

Abbreviations

DINED - <https://www.dined.nl/en>

DINED is an online Anthropometric database tool providing information on using anthropometric data in design.

MT=Mannequin Tool

The Mannequin tool is a user-friendly platform for 3D anthropometric analysis based on 3D scan databases. It allows designers, without the need for an engineering background, to explore 3D body shape variation and create 3D design manikins, e.g. as the basis for a product sizing system.

DINED, MT DINED, MT,

all the words above refer to the mannequin tool in DINED.

Manikin = Mannequin in this project.

AI

Artificial Intelligence

HAI

Human-AI Interaction

AIGC

AI-generated Content

UGC

User-generated Content

PGC

Professor-generated Content

LLM

Large language Model

SD=Stable Diffusion

Stable Diffusion, a text-to-image AI model introduced in 2022, excels at generating intricate images based on textual descriptions.

GUI

Graphical User Interface (GUI)

ComfyUI

ComfyUI is a powerful and modular stable diffusion GUI with a node-based interface that allows users to design and execute advanced stable diffusion pipelines.

LoRA (lora, Lora, LoRa)

LoRA (Low-Rank Adaptation) is a method published in 2021 that offers a way to fine-tune weights in CLIP and UNet models. These models are language models and image de-noisers employed by Stable Diffusion. LoRA enables the modification of Stable Diffusion checkpoint models' output to align with specific concepts or themes, such as art styles, characters, real-life individuals, or objects

Kohya_ss

A GUI for training LoRA model: The Kohya_ss GUI allows users to set the training parameters and run the required CLI commands to train the model.

Preface

Dear Reader,

This is my graduation thesis, written within the context of Industrial Design Engineering at Delft University of Technology, specifically the Master's program in Design for Interaction.

My initial idea for this graduation project was to delve into the realm of human-AI interaction in 3D generation. This interest stemmed from my internship at 51WORLD, a digital twin company, where I became fascinated by editor interaction design in generative AI tools.

Additionally, Dr. Toon Huysmans, the lead of 3D/4D anthropometry at DINED, expressed interest in exploring AI opportunities for the Mannequin Tool (MT) DINED—a 3D mannequin generation anthropometry tool. I wholeheartedly agree that this presents a valuable opportunity to conduct research on human-AI interaction design within the context of a practical tool, while also exploring future possibilities for the MT DINED.

As the research progressed, the team developed a strong interest in text-to-image capabilities and embarked on a series of continuous attempts to combine 3D elements DINED and text-to-image AI. Therefore, I have formulated the title of my graduation thesis as follows:

Design for Diverse Body Shapes with AI: Exploring AIGC Applications for DINED Mannequin Generation

Yunzhao Ma, 2023

Acknowledgments

Because the graduation project is the culmination of my master academic life. It is time to use this opportunity to thank my family, friends, and teachers who have helped shaped my research.

Firstly, I do feel grateful to **my parents** who support me all the time and encourage me to be who I want to be in today's challenging global situation.

Secondly, I want to single out for thanks my two mentors, **Dr. Toon Huysmans** and **Dr. Derek Lomas**, whose help was critical in developing this project. It is the first step that costs troublesome. I proposed my graduation idea with over thirty emails, however, most of them sunk into mailboxes. While Toon and Derek first recognized the value of my graduation project proposal and shed light on my vague idea. Both of them have strong competencies in their fields and at the same time were highly available and friendly. They coached me by sharing their knowledge and tools while giving a good ratio between support and freedom.

Thirdly, I would like to extend my sincere gratitude to **Timon Staal** for his invaluable participation. As an expert in wetsuit tech-design, his expertise in providing user experience insights related to wetsuit body shape design, as well as evaluating my designs with his professional knowledge, proved to be immensely beneficial throughout the entire project. Additionally, I would like to express my gratitude to **Dr. McQuillan, H.L.**, an expert in fashion design research, whose affirmation of this design from the perspective of clothing production empowered me and provided valuable feedback for future recommendations.

Last but certainly not least, I am deeply grateful to my friend **Dinuo Liao**, who shared my passion for generative AI graduation projects. Throughout our journey, Dinuo has been a constant source of inspiration, offering invaluable insights that greatly influenced the development of my project. Furthermore, I extend my sincere appreciation to **Ph.D. Fulin Song** for his invaluable assistance in AI technology. With his expertise and guidance, I was able to successfully integrate the desired functionalities into my project, thereby elevating it to its full potential.

To the many other individuals who have supported and contributed to my project, I am grateful beyond words. Although I cannot mention each of you individually, please know that you hold a special place in my heart.
Thank you all!

Executive Summary

What

In this project, a workflow method integrating generative AI has been developed. This versatile approach can be applied to a variety of body shape-related design tasks. For instance, designers can leverage this method effectively to generate design inspirations, enabling them to establish detailed design themes and goals more efficiently. Building upon this workflow, a new tool named DINED AI has been specifically designed to offer a more user-friendly and guided approach to accomplish these tasks. Although DINED AI is currently in the form of functional prototypes, both the user interface and back-end technology have been individually crafted to provide a realistic demonstration of its capabilities.

Why

With the rapid advancements in large language models, integrating AI technologies has emerged as the prevailing trend in product development. The aim of this project is to explore the potential of creating an AI-powered platform for DINED. DINED is renowned for its specialized expertise in anthropometry, which allows for the generation of realistic data models to aid in the design of ergonomic products. Leveraging this distinctive capability of DINED, and under the guidance of the mentors, this project focuses on the design of clothing, a widely recognized product that is intimately connected to the human body shape's form. Additionally, a streamlined workflow method will be developed to facilitate the design process.

How

The project is centered around leveraging technological advancements by implementing Stable Diffusion, LoRA, and ControlNet as crucial components. Stable Diffusion is a well-established text-to-image algorithm, while the LoRA model effectively refines Stable Diffusion using a limited number of images in the training dataset. On the other hand, ControlNet possesses the unique capability to extract specific control elements from the 3D mannequin generated by DINED, thereby enabling enhanced control over the resulting images. In order to ensure a seamless user experience, the project integrates these aforementioned technologies into a cohesive and streamlined workflow. Each step of the process is accompanied by comprehensive guidance, empowering users to effortlessly navigate the system and fostering a coherent user experience. Furthermore, the project underwent evaluation through expert interviews, further validating its effectiveness and potential impact.

Online Appendix:

[https://www.figma.com/file/5hoD8Rv36mPi1u5kn5Ph7O/80%25--Report-\(Copy\)?type=design&node-id=2878%3A26448&mode=design&t=9fFmpr8UTnAHZrSh-1](https://www.figma.com/file/5hoD8Rv36mPi1u5kn5Ph7O/80%25--Report-(Copy)?type=design&node-id=2878%3A26448&mode=design&t=9fFmpr8UTnAHZrSh-1)

Contents

Chapter 1 Project Overview

- 1.1 Context----- (1-03)
- 1.2 Project Scope And Goal
- 1.3 Design Challenges
- 1.4 Design Approach
- Take Away
- Reference (Chapter 1)----- (1-10)

Chapter 2 AI Research

- 2.1 AI Literature Research----- (2-03)
- 2.2 Specialized AI Model Literature Research
- 2.3 Text-to-image Generation Technology Research In Early 2023
- 2.4 Human-AI Interaction Literature Research
- Conclusion
- Reference (Chapter 2) ----- (2-14)

Chapter 3 Mannequin Tool Research

- 3.1 Product Build-up----- (3-03)
- 3.2 Explorative Interviews
- 3.3 Interview Data Analysis
- 3.4 Context Mapping & Analyzing The Data
- 3.5 Thematic Analysis Results
- Key Insights ----- (3-16)

Chapter 4 Conceptualization

- 4.1 Brainstorming Workshop----- (4-03)
- 4.2 Brainstorming Insights
- 4.3 Insights Combined With User Research
- 4.4 Final Concept
- 4.5 Example Case
- 4.6 Design Goal
- Reference (Chapter 3, 4) ----- (4-14)

Chapter 5 Implementation of Functions

- 5.1 ComfyUI----- (5-03)

5.2 ControNet in ComfyUI

5.3 LoRA Model

5.4 Kohya_ss , A GUI for Training LoRA Models

5.5 Case1 : Efficiently Generate Inspired Visuals

5.6 Case2 : Efficiently Generate Inspired Visuals

Reference (Chapter 5) -----(5-16)

Chapter 6 DINED AI Tool Interaction Design

UX & UI Design----- (6-03)

6.1 STEP_1_Foundation Type Selection

6.2 STEP_2_Object Effect Settings

6.3 STEP_3_Human Factors Settings

6.4 STEP_4_ Image Layout (Optional)

6.5 STEP_5_Final Effect

Take Away -----(6-12)

Chapter 7 Research Insights

7.1 Insights on ControlNet----- (7-03)

7.2 Insights on LoRA Model Training

7.3 Obese Body Shape & Base Model

7.4 Overall Comparison

Take Away

Reference (Chapter 6, 7)----- (7-14)

Chapter 8 Evaluation

8.1 Evaluating With Experts----- (8-03)

8.2 Expert Interview Brief 1

8.3 Expert Interview Brief 2

Conclusion

Take Away----- (8-10)

Chapter 9 Limitations, Suggestions, and Future Work

9.1 Limitations & Recommendations----- (9-03)

9.2 Future Work

Take Away -----(9-16)

Chapter 10 Reflection

10.1 Reflection----- (10-03)

10.2 Supplement

Reference (Chapter 8, 9, 10)----- (10-07)

01

Project Overview



Chapter 1 This chapter is a summary of the project, including its context, project scope & goal, design challenges, and design approach. It provides an overview of the project and helps readers understand what the project is all about.

Chapter Overview

1.1 Context

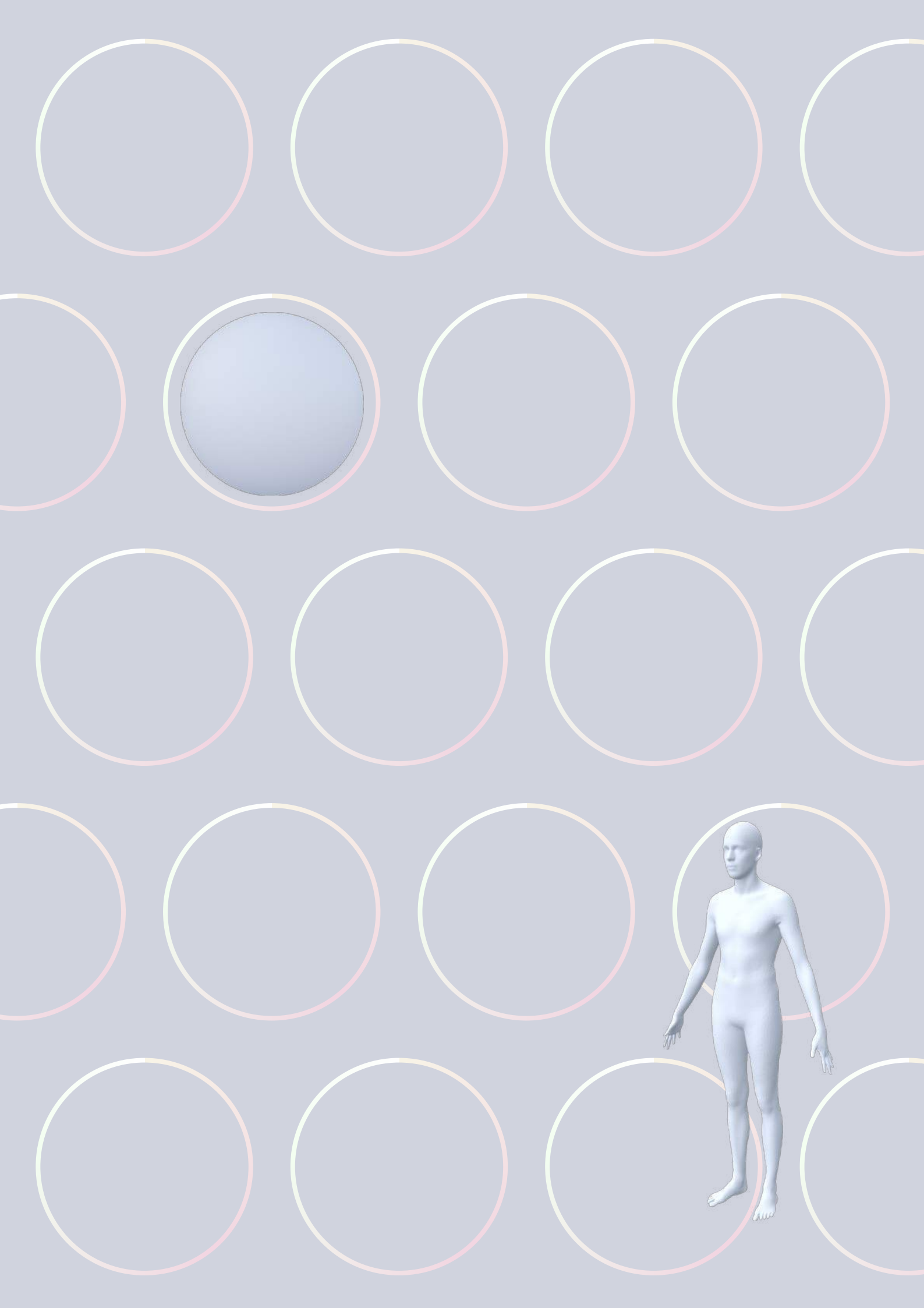
1.2 Project Scope And Goal

1.3 Design Challenges

1.4 Design Approach

Take away

Reference



1.1 Context

In the current context of accelerated integration of the digital and physical worlds, AIGC is quietly leading a profound change, reshaping or even overturning the production and consumption modes of digital content. It will greatly enrich people's digital life and is an indispensable support force for the future development of a new era of digital civilization. While DINED is searching for brand new possibilities in AI age.

Human-AI Interaction Design

AIGC models have the potential to fundamentally change the way we create high-fidelity digital content but are often hard to control. With the increasing diffusion and capacity of complex algorithms, there is a need to better understand how AI interfaces should be designed to be able to effectively use and control the system [1].

Generative AI

The recent progress in generative AI has led to surprisingly good text-to-content generative models:

text-to-image models, like DALL-E, Midjourney, or Stable Diffusion...

text-to-3D models, like Nvidia Magic3D and Google DreamFusion...

text-to-manikin models, A-NeRF and H-NeRF

text-to-motion models[2].

Those AIGC models are raising concerns about Human-AI interaction:

“How will AI assist people in the future content creation field?”

“How should the interaction design develop to let products with text-to-content AI algorithm provide a user-friendly experience?”

DINED Tool

The DINED / Mannequin tool is a user-friendly platform for 3D anthropometric analysis based on 3D scan databases. It allows designers, without the need for an engineering background, to explore 3D body shape variation and create 3D design manikins, e.g. as the basis for a product sizing system [3].

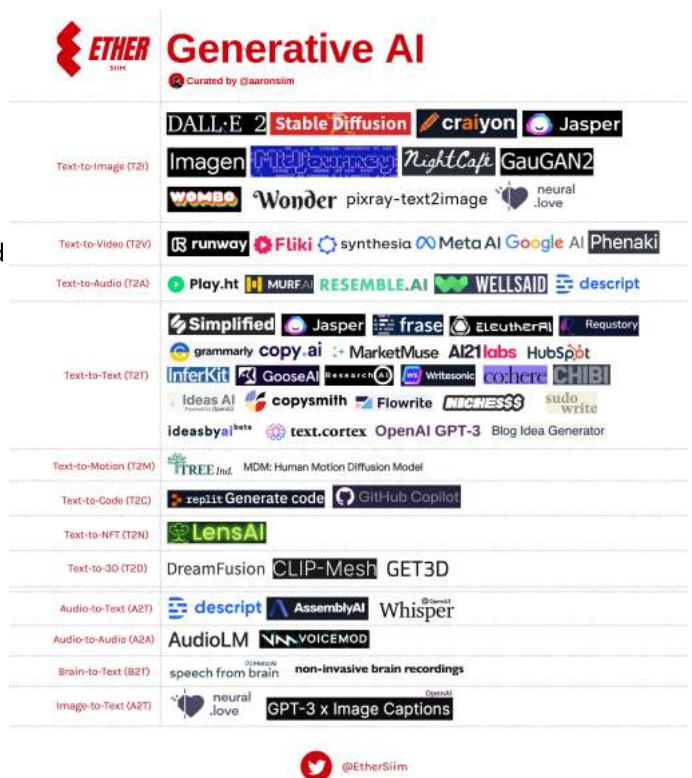


Figure 1.1-1 Overview of Popular Generative AI Products
curated by @aaronSiim

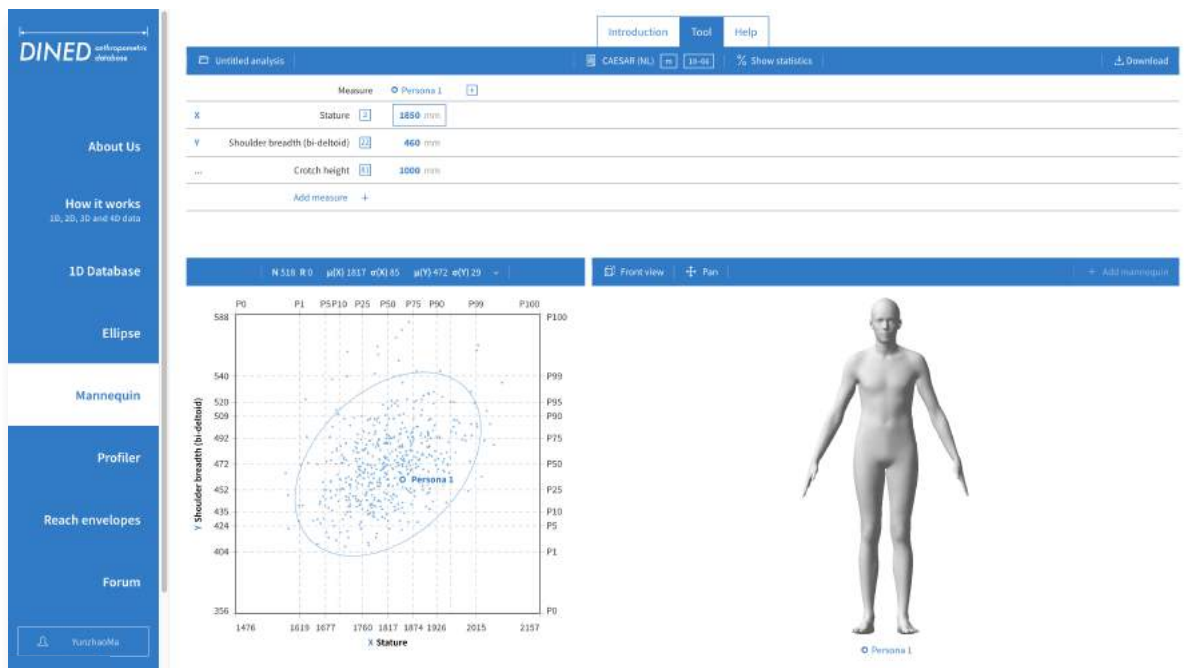


Figure 1.1-2 Screenshot of the Mannequin Tool

1.1 Reference

- [1] Amershi, S., Weld, D., Vorvoreanu, M., Fournay, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019, May). Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1-13).
- [2] Roberts, J., Banburski-Fahey, A., & Lanier, J. (2022). Steps towards prompt-based creation of virtual worlds. *arXiv preprint arXiv:2211.05875*.
- [3] DINED. (n.d.). <https://www.dined.nl/en/mannequin/introduction>

1.2 Project Scope And Goal

Project Scope

The research landscape, which consists of human-AI interaction and the DINED/Mannequin Tool, has been unfolded in the previous section, and thereof the research topic is identified in endowing the DINED with AI generative ability and extracting human-AI interaction insights for the design practice.

In order for DINED to thrive in the age of AI, it is imperative to remain at the forefront of technological advancements and continually explore new technologies. This commitment allows us to deliver superior services to our users.

Furthermore, due to its ease of implementation, Large Language Models (LLMs) involved in generative AI is a fascinating topic that has gained a lot of attention recently. However, it is important to note that such AI models have their limitations. One concern is the potential for the models to generate misleading or biased content. LLMs learn from the data they are trained on, which can include biased or inaccurate information. This can lead to the generation of content that reflects those biases or inaccuracies [1]. Thus, there are a lot of unexpected questions happening in design practice.

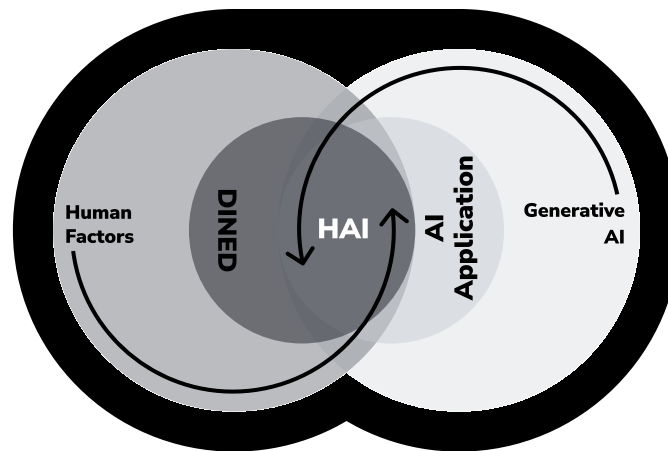


Figure 1.2-1 Project Scope (HAI: Human-AI Interaction)

There are two goals in my graduation project. The first goal is to explore human-AI interactions of anthropometry body shape generation & generative AI models and figure out a AIGC solution for MT. The second goal is to formulate guidelines for AIGC products based on the insights found in the exploring process.

Herewith, exploratory research questions are formulated to navigate the process of exploration in the fuzzy front end (Sander & Stapper, 2012 [2]) of this project before defining what are the most advisable method to update the MT to AIGC products.

1. Research Questions in the First Goal:

Human-AI interactions of anthropometry body shape generation & generative AI models and figure out a AIGC solution for MT

1. Are Text-to-Image, Text-to-Motion, Image-to-Manikin, Action-to-Motion, and other kinds of generative AI models could and should be involved in MT? Why?
2. If the answer is yes, then how to apply the generative model to MT?
3. Are there any other potential AI capabilities that could be used to enhance the UX of MT? How to integrate AI capabilities as a whole smooth UX into MT?

2. Research Questions in the Second Goal:

Formulating human-AI interaction design guidelines for AIGC products

1. Is there anything valuable for future human-AI interaction design in AIGC products?
2. How can the AI model be trained or used to generate the most accurate results?
3. What steps can be taken to ensure that the AI model is not introducing any body shape bias?
4. How can the AI model be evaluated to ensure that it is meeting the given objectives?
5. What strategies can be used to ensure that AI models are delivering the best possible user experience?

1.2 Reference

- [1] Mearian, L. (2023, May 30). What are LLMs, and how are they used in generative AI? Computerworld. <https://www.computerworld.com/article/3697649/what-are-large-language-models-and-how-are-they-used-in-generative-ai.html>
- [2] Sanders, E. B. N., & Stappers, P. J. (2012). Convivial toolbox: Generative research for the front end of design. *Bis*.

1.3 Design Challenges

The challenge of human-AI interaction design lies in finding the best balance between the two. AI is a powerful tool that can automate processes and provide useful insights, but it should not replace human interaction or decision-making. AI should be used to augment and improve human decision-making, rather than replace it.

The success of human-AI interaction design hinges on creating an environment where both human and AI can work together to achieve the desired outcome. This involves understanding the strengths and limitations of both parties, and creating an interface that can be easily used by both. Designers should also strive to create systems that are transparent and understandable so that users can trust the decisions made by the AI.

The goal of human-AI interaction design is to create a symbiotic relationship between humans and AI. By carefully balancing the strengths of both, designers can create powerful and versatile systems that can help to make better decisions and improve outcomes.

Human-AI interaction design challenges and mitigation considerations (Babak Abedin et., 2022)

AI User Interface Design

1. Which challenges come with AI technology regarding the design of user interfaces and hence various modes of human-AI interaction ?
2. How do interfaces need to be designed in order to increase trust and technology acceptance ?

Human-AI Conversations and Collaboration

1. How should communication between AI and humans be designed in order to avoid misunderstandings on both ends ?
2. How can AI be designed in order for it to adapt to individual characteristics of humans?

Explainability, Accountability, Ethics, Fairness and Bias

1. How can explainability be designed into the process from AI development by experts (e.g., engineers) to AI usage by laymen?
2. To which extent may we need to expand or re-think the term 'explainability' in the context human-AI interaction?
3. What are AI system design considerations for accountability, ethics, fairness and bias of AI systems in operation in human society

AI Agency and Human Interaction with Agentic AI

1. To which degree can and should agency be designed into AI?
2. How does the notion of agentic AI influence the design of human-AI interaction ?

1.3 Reference

[1] Abedin, B. (2022). Managing the tension between opposing effects of explainability of artificial intelligence: a contingency theory perspective. *Internet Research*, 32(2), 425-453.

1.4 Design Approach

Here I want to introduce the design methodologies applied throughout this project. Overall the model is a variant of the **Double Diamond** adapted from the divergence-convergence model proposed in 1996 by Hungarian-American linguist Béla H. Bánáthy [1] [2]. Based on the 4 design phases (Discover, Define, Develop, and Deliver) in DD, I established **circular unit** for my human-AI interaction research workflow. Designed according to two parts of this project scope and goal, the circular unit helped me better reflect on the design process by making a clear distinction between doing and thinking.

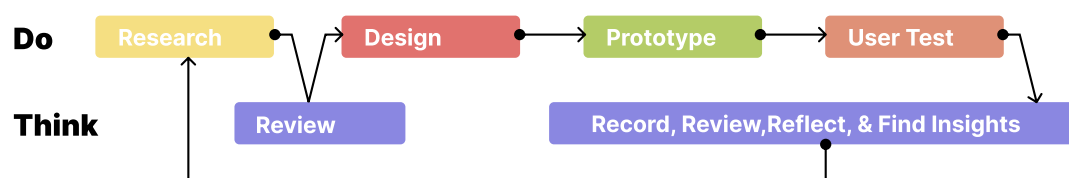


Figure 1.4-1 Circular Unit According to the scope and goal, there are two parallel parts in my project: The first part is about "Do" including research (yellow tag), design (red tag), prototype (green tag), programming (blue tag), and user test (orange tag). The second part is about "Think" consisting of "record", "review", "reflect", and "find insights". All of them are purple tags. The circular unit is "Research (yellow tag)+Review (purple tag)+Design (red tag)+ Prototype (green tag)+User Test (orange tag)+Record & Reflect (purple tag)".

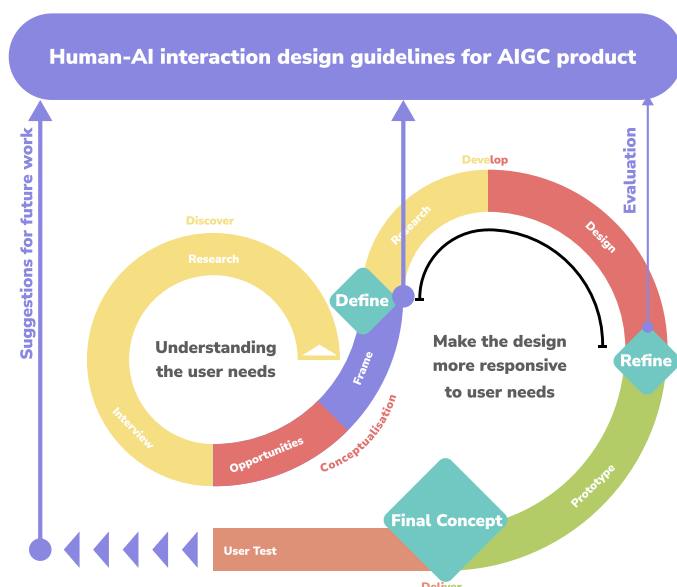


Figure 1.4-2 Overview Methodology after bringing circular units into Double Diamond, there is one more outcome of high level guidance besides the design delivery in DD. The first unit focus on understanding the user needs. The second unit is to make the human-AI interaction design more responsive to user needs.

Supplementary :

In *Discover* phase, the context and design challenges were first explored by **literature studies**. Through **interviews** with users of the Mannequin Tool, I was able to understand the user needs about the MT and expectations of AIGC ability. By conducting interviews with designer users and observations on how they generate 3D manikin model online, I got a better understanding of the inspirational process of creators. In *Define* phase, I set my **design goal (DG)** covering the required interaction qualities to encourage ideation.

Involve Pattern Language Theory (PLT) in AI literature Research

PLT is a methodology that utilizes patterns and pattern languages to deal with complex systems. On one hand, this methodology allows knowledge to be easily captured and communicated (Borchers, 2000 [3] ; Van Welie & Van der Veer, 2003 [4] ; Erickson, 2000 [5];). On the other, it can leverage consistency and create an overview of wicked problems (Finidori, 2014 [6]) or complex systems. In this project, it is a great challenge for none AI developer like me to closely follow and fully understand the frontier AI algorithms especially in this AI blooming year of 2023. Also, it is crucial to discuss about AI algorithms with user participants during research. Thus, I make AI tags and cards with PLT to assist me to construct related AI ability map and communicate with user participants.

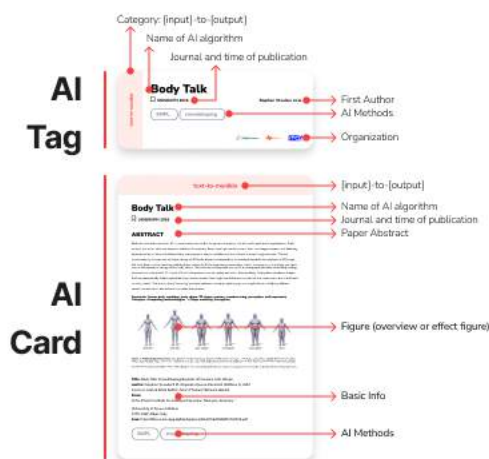


Figure 1.4-3 AI Tag & AI Card with marked pattern languages

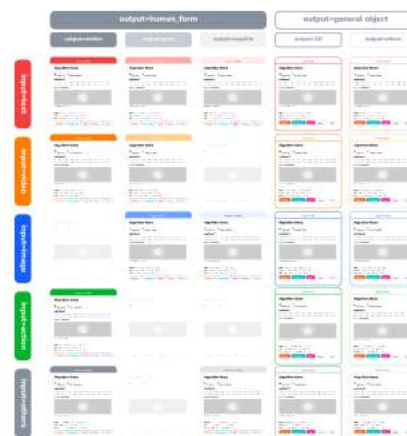


Figure 1.4-4 Set color and style patterns to AI Cards according to the categories of AI input and output

1.4 Reference

- [1] Banathy, Bela H. (1996). *Designing Social Systems in a Changing World*. Springer US. p. XV, 372. ISBN 978-0-306-45251-2.
- [2] Möller, Ola (9 January 2015). "The Double Diamond". *MethodKit Stories*. Retrieved 3 September 2019.
- [3] Borchers, J. O. (2000, August). A pattern approach to interaction design. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques* (pp. 369-378).
- [4] Van Welie, M., & Van der Veer, G. C. (2003, September). Pattern languages in interaction design: Structure and organization. In *Proceedings of interact* (Vol. 3, pp. 1-5).
- [5] Erickson, T. (2000, August). Lingua Francas for design: sacred places and pattern languages. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques* (pp. 357-368).
- [6] Finidori, H. (2016). Patterns that connect: exploring the potential of patterns and pattern languages in systemic interventions towards realizing sustainable futures. In *Proceedings of the 60th Annual Meeting of the ISSS-2016 Boulder, CO, USA* (Vol. 1, No. 1).

Take Away

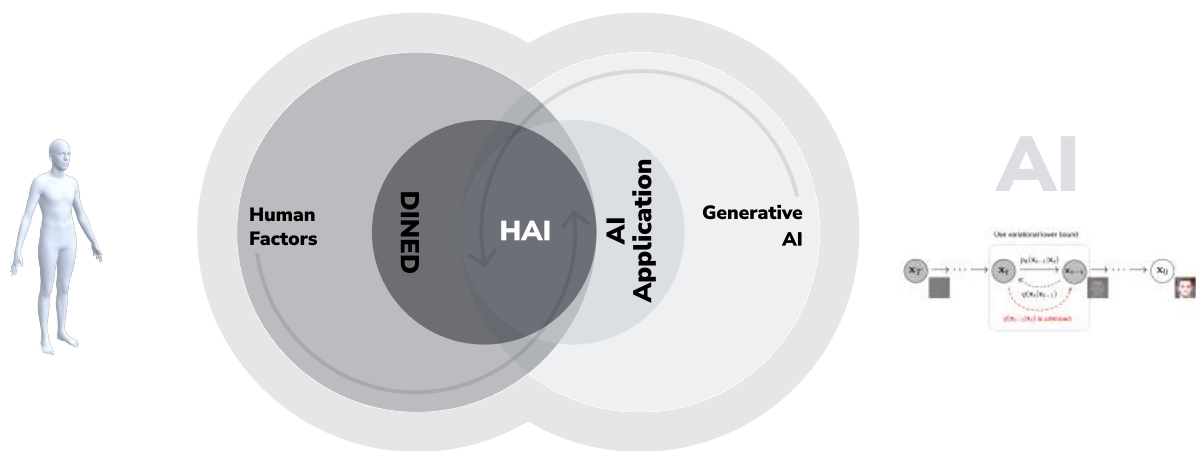


Figure 1.4-1 Project Scope Concept

This project highlights concerns about the control and design of AIGC models and explores the potential of generative AI models in various content creation fields. The DINED/Mannequin tool, a user-friendly platform for 3D anthropometric analysis, is introduced.

The project's scope and goals include endowing DINED with AI generative ability, extracting human-AI interaction insights, and formulating guidelines for AIGC products.

The research questions focus on exploring human-AI interactions, applying generative models to content creation, enhancing user experience, and addressing design challenges related to AI technology, trust, communication, ethics, fairness, and agency.

In addition, this project use a variant of the **Double Diamond** to guide this project.

Reference (12)

Chapter 1

1.1 Reference

- [1] Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019, May). Guidelines for human-AI interaction. In Proceedings of the 2019 chi conference on human factors in computing systems (pp. 1-13).
- [2] Roberts, J., Banburski-Fahey, A., & Lanier, J. (2022). Steps towards prompt-based creation of virtual worlds. arXiv preprint arXiv:2211.05875.
- [3] DINED. (n.d.). <https://www.dined.nl/en/mannequin/introduction>

1.2 Reference

- [1] Mearian, L. (2023, May 30). What are LLMs, and how are they used in generative AI? Computerworld. <https://www.computerworld.com/article/3697649/what-are-large-language-models-and-how-are-they-used-in-generative-ai.html>
- [2] Sanders, E. B. N., & Stappers, P. J. (2012). Convivial toolbox: Generative research for the front end of design. Bis.

1.3 Reference

- [1] Abedin, B. (2022). Managing the tension between opposing effects of explainability of artificial intelligence: a contingency theory perspective. Internet Research, 32(2), 425-453.

1.4 Reference

- [1] Banathy, Bela H. (1996). Designing Social Systems in a Changing World. Springer US. p. XV, 372. ISBN 978-0-306-45251-2.
- [2] Möller, Ola (9 January 2015). "The Double Diamond". MethodKit Stories. Retrieved 3 September 2019.
- [3] Borchers, J. O. (2000, August). A pattern approach to interaction design. In Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques (pp. 369-378).
- [4] Van Welie, M., & Van der Veer, G. C. (2003, September). Pattern languages in interaction design: Structure and organization. In Proceedings of interact (Vol. 3, pp. 1-5).
- [5] Erickson, T. (2000, August). Lingua Francas for design: sacred places and pattern languages. In Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques (pp. 357-368).
- [6] Finidori, H. (2016). Patterns that connect: exploring the potential of patterns and pattern languages in systemic interventions towards realizing sustainable futures. In Proceedings of the 60th Annual Meeting of the ISSS-2016 Boulder, CO, USA (Vol. 1, No. 1).

02

AI Research



Chapter 2 focuses on AI research, encompassing two primary areas: AI model research and human-AI research. This chapter delves into the advancements, application methods of both domains, shedding light on their significance in shaping the theoretical basis of this project.

Chapter Overview

2.1 AI Literature Research

2.2 Specialized AI Model Literature Research

2.3 Text-to-image Generation Technology Research In Early 2023

2.4 Human-AI Interaction Literature Research

Conclusion

Reference

2.1 AI Literature Research

Related generative AI models

Action category based human motion generation also draws considerable interests by resorting to a diverse range of learning strategies, including GANs, VAEs , Transformers.

Here is a quick summary of the difference between GAN, VAE, and flow-based generative models:

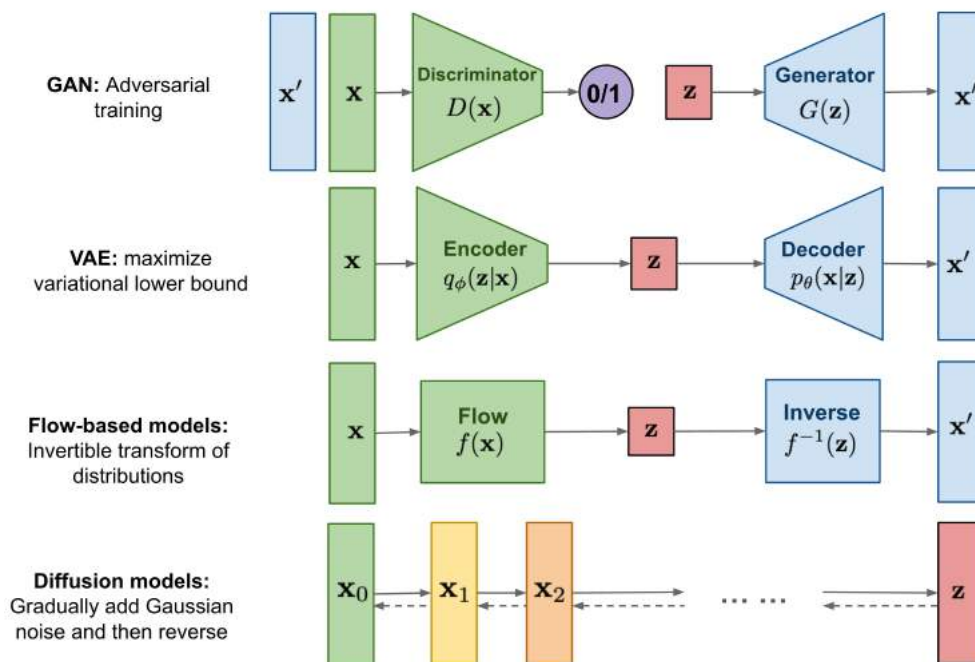


Figure 2.1-1 Overview of different types of generative models [1].
Please See Appendix C2.1

1. Generative Adversarial Networks: GAN provides a smart solution to model the data generation, an unsupervised learning problem, as a supervised one. The discriminator model learns to distinguish the real data from the fake samples that are produced by the generator model. Two models are trained as they are playing a minimax game.

2. Variational Autoencoders: VAE in explicitly optimizes the log-likelihood of the data by maximizing the evidence lower bound (ELBO).

3. Flow-based Generative Models: A flow-based generative model is constructed by a sequence of invertible transformations. Unlike other two, the model explicitly learns the data distribution $p(x)$ and therefore the loss function is simply the negative log-likelihood.

4. Diffusion Models: In machine learning, diffusion models, also known as diffusion probabilistic models, are a class of latent variable models. They are Markov chains trained using variational inference. The goal of diffusion models is to learn the latent structure of a dataset by modeling the way in which data points diffuse through the latent space.

Applications & Examples

GAN



Figure 2.1-2 When a computer puts Nicolas Cage's face on Elon Musk's head, it may not line up the face and the head correctly.[2]

VQ-VAE



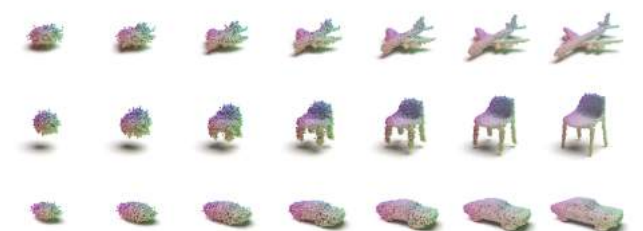
Figure 2.1-3 DALL_E 1

Diffusion



Figure 2.1-4 DALL_E 2

Flow-based Generative Models



<https://www.guandaoyang.com/PointFlow/>

Figure 2.1-5 PointFlow

GAN

High-fidelity natural image synthesis
Data augmentation tasks
Improving image compression
Image inpainting or hole filling
Generating synthetic data
Image super-resolution
Image translation
3D object generation
Steganography
Video prediction

VAE

Generative modeling: VAEs can generate new data that is similar to the training data, and can also modify existing data in specific directions.

Compact representation: VAEs learn to generate compact representations and reconstruct their inputs well.

Medical applications: VAEs are also used in medical applications to detect ventilator-associated events (VAEs) in patients.

Diffusion

Image synthesis
Video generation
Molecule design
Efficient sampling
Improved likelihood estimation
Handling data with special structures

Flow-based Generative Models

Flow-based generative models have been applied on a variety of modeling tasks, including:

Audio generation
Image generation
Molecular graph generation
Point-cloud modeling
Video generation
Lossy image compression

2.1 Reference

- [1] Weng, L. (2021, July 11). What are Diffusion Models? Lil'Log. <https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>
[2] Lyu, S. (2019, June 26). Detecting deepfakes by looking closely reveals a way to protect against them. phys.org. <https://phys.org/news/2019-06-deepfakes-reveals.html>

2.2 Specialized AI Model Literature Research

AI Tags & AI Cards : The author employs **Pattern Language Theory (PLT)** to organize AI knowledge, which is a methodology that utilizes patterns and pattern languages to deal with complex systems.

Figure 2.2-1 Legend of AI Cards Clustering the literature according to the AI model input and output content categories



AI Cards



Please See Appendix C2.2

2.3 Text-to-image Generation Technology Research In Early 2023

Open Source AI Model

Stable Diffusion

Stable Diffusion, a text-to-image AI model introduced in 2022 [1], excels at generating intricate images based on textual descriptions [1]. Moreover, it exhibits its usefulness in various other tasks, including inpainting, outpainting, and creating image-to-image translations guided by textual prompts [1]. With extensive training on billions of images, Stable Diffusion's results are comparable to those achieved by DALL-E 2 and MidJourney [2].

The model offers multiple applications, ranging from generating images solely from text (text2img) to modifying existing images using both text and image inputs [3]. It demonstrates its versatility by accommodating diverse styles, with the availability of high-quality checkpoints fine-tuned specifically for particular styles [4].

Stable Diffusion, an open-source machine learning model, can be accessed through its official release on GitHub [5]. Over time, the model has undergone several iterations, with the latest version being Stable Diffusion 2.1, which is recognized as a significant improvement compared to its predecessor, Stable Diffusion 2.0. The earlier version received criticism for producing subpar results, particularly in the depiction of people [6].

To utilize Stable Diffusion, users have multiple options, including web-based applications and installation choices. One notable example is DreamStudio by Stability AI, which offers a user-friendly interface for generating images based on textual prompts [7]. However, it is important to bear in mind that web versions might have limitations in terms of available options and output quality, in contrast to installed versions [7].

In conclusion, Stable Diffusion is a powerful text-to-image generation model that has been widely adopted in the AI community. It offers versatility in terms of image generation and alteration, and its open-source nature allows for continuous improvements and adaptations to specific styles and applications.

2.3 Stable Diffusion Reference

- [1] Wikipedia contributors. (2023). Stable Diffusion. Wikipedia. https://en.wikipedia.org/wiki/Stable_Diffusion
- [2] Lewis, N. (2022b). How to Run Stable Diffusion on Your PC to Generate AI Images. How-To Geek. <https://www.howtogeek.com/830179/how-to-run-stable-diffusion-on-your-pc-to-generate-ai-images/>
- [3] Wuxiaojun. (n.d.). Diagram diffusion model: Understanding stable diffusion from the perspective of model. Zhihu column. <https://zhuanlan.zhihu.com/p/581186398>
- [4] Effective and efficient diffusion. (n.d.). https://huggingface.co/docs/diffusers/stable_diffusion
- [5] Lewis, N. (2022c). How to Run Stable Diffusion on Your PC to Generate AI Images. How-To Geek. <https://www.howtogeek.com/830179/how-to-run-stable-diffusion-on-your-pc-to-generate-ai-images/>
- [6] Ma, Y. (2023). The 10 Best Stable Diffusion Models by Popularity (SD Models Explained). AiTuts. <https://aituts.com/models/>
- [7] Boehman, C. (2023). Midjourney vs. DALL-E vs. Stable Diffusion: Which Is Better? MUO. <https://www.makeuseof.com/midjourney-vs-dalle-vs-stable-diffusion/>

Relatively Mature Products

DALL_E2

DALL-E 2, an artificial intelligence model developed by OpenAI, has been specifically designed to generate high-resolution, lifelike images based on textual prompts. Serving as the successor to the original DALL-E model introduced in January 2021 [1], DALL-E 2 was announced in April 2022 with notable enhancements. This new iteration excels in producing images with four times the resolution of its predecessor while also offering increased customization options, including various styles and outpainting capabilities [2].

The functioning of DALL-E 2 relies on the integration of natural language processing (NLP), large language models (LLMs), and diffusion processing. It utilizes a subset of the GPT-3 LLM, which comprises an impressive 12 billion parameters specifically optimized for image generation. Through a transformer neural network, the model adeptly establishes and comprehends connections between different concepts, enabling its impressive image generation capabilities [2].

In contrast to the original DALL-E model that utilized a Discrete Variational Auto-Encoder (dVAE) for image generation, DALL-E 2 takes a different approach by employing a diffusion model. This updated model incorporates data from the CLIP model, resulting in the generation of higher quality images [2]. The diffusion model operates by selecting a sequence of freely positioned dots and modifying the sequence as it progressively captures visual aspects, effectively imitating the image generation process [3].

DALL-E 2 introduces a range of customization options, enabling users to tailor images according to different styles, such as pixel art or oil painting. Additionally, it introduces the concept of outpainting, which allows users to extend an original image and create an image that seamlessly complements it [2].

Regarding commercial use, DALL-E 2 offers users the ability to utilize the generated images for various business purposes, such as printing, trading, or licensing. It is important to note that a watermark is automatically added to the corner of the image to ensure proper source credit is given by users [3].

While OpenAI has not released the source code for either version of DALL-E, they have made DALL-E 2 available as an API starting from early November 2022. This API allows developers to seamlessly integrate the DALL-E 2 model into their own applications. Notably, several companies, including Microsoft, CALA, and Mixtiles, have been early adopters of the DALL-E 2 API, incorporating its capabilities into their respective offerings [1].

Looking ahead, the future of DALL-E 2 holds potential for generating new corporate aesthetics, replacing generic stock portraits with captivating visual content, and expanding its applications within the realms of creative marketing and branding. These advancements have the potential to revolutionize visual representation in various industries [3].

2.3 DALL_E2 Reference

[1] Wikipedia contributors. (2023a). DALL-E. Wikipedia. <https://en.wikipedia.org/wiki/DALL-E>

[2] Kerner, S. M. (2023). Dall-E. Enterprise AI. <https://www.techtarget.com/searchenterpriseai/definition/Dall-E>

[3] What is DALL-E 2? Use Cases and Examples. (2023, January 2). Cubix. <https://www.cubix.co/blog/what-is-dall-e-2>

Chapter 2

Midjourney

Midjourney is an innovative generative AI program and service offered by the independent research lab, Midjourney, Inc, headquartered in San Francisco. Currently in its open beta phase, Midjourney was launched on July 12, 2022 [1]. The distinguishing characteristic of Midjourney lies in its unique dream-like and artistic style, setting it apart as more of a painting tool than a conventional photo-based generator. This distinctive approach makes it especially well-suited for individuals involved in science-fiction literature or those seeking to evoke a gothic ambiance in their artwork [2].

Midjourney, a text-to-picture AI tool, can be accessed via a Discord bot on their official server. Users utilize the /imagine command and enter a prompt to generate four images, allowing them to choose which ones to upscale [1].

Advantages of Midjourney as a text-to-picture AI tool include:

- High-quality images: Midjourney produces well-structured, defined, and realistic images [3].
- High image resolution: It generates images with resolutions up to 1,792 x 1,024 pixels [3].
- Freemium model: Users can try the service with a limited number of free image creations before opting for a paid plan [3].
- User-friendly platform: Midjourney operates through Discord, making it accessible even to non-coders [3].
- Customization options: Users can fine-tune their image creations using various commands and parameters [3].
- Active Community: The Midjourney Discord server fosters an engaged community for sharing creations, asking questions, and receiving support from other users and the Midjourney team [3].
- Continuous development: The Midjourney team consistently enhances and expands the AI capabilities, ensuring the tool remains relevant in the evolving AI landscape [3].

Midjourney offers a distinctive feature that allows users to upload their own images and have the platform generate descriptive tags. These tags can then be used as prompts to create entirely new images. By using the /describe command and uploading an image, Midjourney provides four prompts based on its interpretation of the image content [4].

To join Midjourney, users have the option to sign up through the company's website or by downloading and using the dedicated Discord app. Upon registration, users receive a link to the Midjourney Discord server, where they can interact with the AI bot using simple commands [5]. While the platform was in beta, access was available without a subscription. However, to enjoy full access to all features, a subscription plan is now required.

2.3 Midjourney Reference

[1] Wikipedia contributors. (2023a). DALL-E. Wikipedia. <https://en.wikipedia.org/wiki/DALL-E>

[2] Hughes, A. (2023). Midjourney: The gothic AI image generator challenging the art industry. www.sciencefocus.com. <https://www.sciencefocus.com/future-technology/midjourney/>

[3] Urrutia, D. (2023). What is Midjourney | Definition, advantages and how to use it. Arimetrics. <https://www.arimetrics.com/en/digital-glossary/midjourney>

[4] Boehman, C. (2023b). Midjourney vs. DALL-E vs. Stable Diffusion: Which Is Better? MUO. <https://www.makeuseof.com/midjourney-vs-dalle-vs-stable-diffusion/>

[5] Melekhin, K. (2023, March 15). What is Midjourney? A Comprehensive Guide. HostAdvice. <https://hk.hostadvice.com/blog/ai/what-is-midjourney/>

Novel_AI

Novel AI is a versatile artificial intelligence (AI) platform that supports users in various creative writing endeavors, offering AI-assisted authorship, storytelling, virtual companionship, and a Natural Language Processing (NLP) playground [1]. Specifically designed for novelists, fiction writers, and individuals seeking a virtual space to nurture their imagination [2], Novel AI's AI models are trained on real literature, enabling them to comprehend and emulate diverse writing styles, perspectives, and themes [1].

Key features of Novel AI include:

- **AI-Assisted Authorship:** The platform generates writing that resembles human-like quality based on user input, empowering users to produce compelling literature regardless of their writing abilities [1].
- **Storytelling:** Novel AI adapts to users' input, preserving their unique perspective and style, allowing them to shape narratives according to their vision [1].
- **Virtual Companionship:** Users can engage in simulated conversations with AI-generated characters, providing a personalized and immersive experience [1].
- **Natural Language Processing Playground:** The platform serves as a playground for exploring NLP capabilities, enabling users to experiment with AI-generated text, expand their creativity, and enhance their writing skills [1].

Novel AI goes beyond writing assistance and extends its offerings to include a text-to-image AI art generator powered by Stable Diffusion, enabling users to bring their stories to life visually [3]. Additionally, the platform provides tools for transforming text into images and speech [4].

To access Novel AI's features, users need to subscribe to a monthly plan, with different tiers offering various benefits [5]. The platform is conveniently accessible via a web browser, ensuring usability across mobile devices [5].

In comparison to other AI image generators, Novel AI distinguishes itself by its unique ability to generate complete stories from a single word input and its integrated illustration capabilities within the app [6]. However, users are encouraged to explore alternative options and consider their specific requirements before making a decision to utilize the platform [6].

2.3 Novel_AI Reference

[1] Goodwin, R. (2023, June 6). What is Novel AI? Features & How It Works Detailed. Know Your Mobile. <https://www.knowyourmobile.com/ai/novel-ai-features/>

[2] NovelAI - The AI Storyteller. (n.d.). NovelAI. <https://novelai.net/>

[3] Visimens-Technology. (2022, October 14). What is a novel ai and how to use it - visimens technology. visimens technology. <https://visimens-technology.com/what-is-a-novel-ai-and-how-to-use-it/>

[4] Paiva, R. (2023, March 23). What is NovelAI? Review - Robot Artificial. Robot Artificial. <https://www.robotartificial.com/what-is-novelai-review/>

[5] The_Human_Empire. (n.d.). Reddit - Dive into anything. https://www.reddit.com/r/NovelAi/comments/oe16kf/what_exactly_is_novel_ai/

[6] Chainanon. (2023). NovelAI Review: How good is the AI storyteller really? AiTuts. <https://aituts.com/ultimate-novel-ai-review/>

2.4 Human-AI Interaction Literature Research

Stable Diffusion

Designing and Managing Human-AI Interactions

The framework contained four research challenges across two levels of considerations.

With the increasing diffusion and capacity of complex algorithms, there is a need to better understand how AI interfaces should be designed to be able to effectively use and control the system. This need is rooted in AI's complexity and functional capability, its many shapes and forms, and the varying levels of user expertise required. [1]

While the intention is to make AI systems' user interfaces more effective and easier to use, perceptions about the unreliability of AI and that its interface is merely cosmetic has created disagreements about how to design guidelines and principles for AI user interface alternatives.

Interestingly, our analysis of the submissions to the Special Issue revealed a greater attention to two of those challenges, human-AI conversations and collaboration, as well as explainability, accountability, fairness and bias. Less attention received AI user interface design and AI agency and human-interaction with agentic AI.

The author believes moving forward that scholars need to further theorize agentic AI, examine in what ways interactions with agentic AI would be similar or different from a non-agentic AI and information systems, and further study the implications and specifications of AI's user-interface design in interactions with humans.

Toward General Design Principles for Generative AI Applications

Given the surge in popularity of generative applications, there is a pressing need for guidance on designing these applications to promote both productivity and safety, while aligning with human-centered AI values. Their principles aim to achieve three main objectives:

1. Providing designers of generative AI applications with a language to address unique issues pertaining to generative AI.
2. Offering strategies and guidance to assist designers in making crucial decisions regarding user interactions with generative AI applications.
3. Raising designers' awareness about the potential harms associated with generative AI applications.

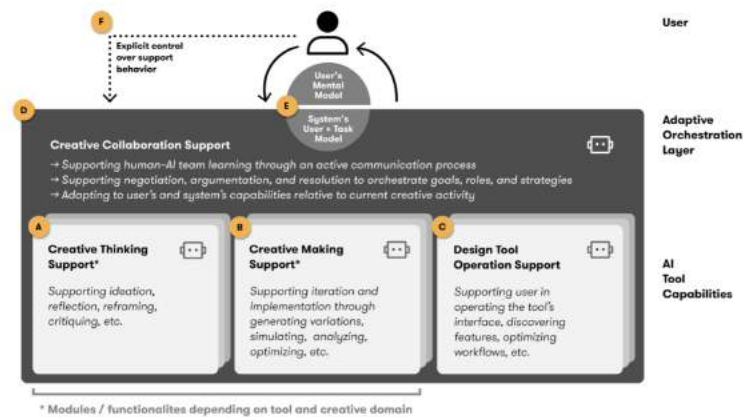
Designers should recognize that the system might generate multiple outputs for a given query, and these outputs may contain flaws or imperfections. It is crucial for users to actively collaborate with the system, utilizing various controls, in order to produce desired artifacts. Additionally, designers should include adequate controls that empower users to explore a wide range of possibilities, thus fostering inspiration and creativity.

Furthermore, designers should enable users to seek clarification on specific model outputs. Users should have the ability to inquire about the model's confidence in a given output, identify portions of the output that may require human review or revision, learn how to adjust or modify the input to influence output properties, and explore alternative options or alternatives.

The authors present a set of seven design principles for generative AI applications. These principles are grounded in an environment of generative variability, the key characteristics of which are that a generative AI application will generate artifacts as outputs, and those outputs may be varied in nature.

Interactive Prototype Design Ideas

Figure 2.4-1 Diagram illustrating the schematic framework for supporting co-creative human-AI team learning comprising (AC) core capabilities supporting creative processes and tool usage, and (D) adaptive support layer for orchestrating creative collaboration between user and system by building (E) shared mental models. (F) Users have explicit control over the system's support behavior [3]



Team Learning as a Lens for Designing Human-AI Co-Creative Systems

This paper delves into a research agenda that focuses on understanding how collaboration between human-AI teams can take place in co-creative task domains. The research methodology involves conducting formative think-aloud studies to gain valuable insights into how professional users learn to collaborate with existing generative AI tools. The objective is to identify support functionalities, learning strategies, and the challenges faced when collaboration learning support is not specifically provided. Additionally, the research includes simulations of an AI support system where human experts assist users to gather insights for designing a team learning support system. The primary goal of this research is to identify specific situations and opportunities that facilitate effective communication and the development of shared mental models among human-AI teams. Furthermore, data will be collected to understand users' general attitudes towards AI-driven co-creation and their collaboration preferences concerning agency, control, and team roles. The findings from this research will provide valuable input for designing novel interfaces and interactions that support human-AI co-creation. This will be achieved through an iterative process of prototyping intelligent user interfaces.

The researchers also aim to evaluate the effectiveness of the developed prototypes in facilitating co-creative collaborations between humans and AI. In addition, they seek to identify learning goals and develop guidelines that will inform the design of future interfaces for generative AI tools, with a specific focus on enhancing human-AI collaboration. The research raises thought-provoking questions for further discussion, such as the suitability of team learning as an approach for all users and creative domains, the potential team learning goals in co-creative systems, and the modalities and communication styles that support non-disruptive team learning during flow-state co-creative activities.

2.4 Reference

[1] Abedin, B., Meske, C., Junglas, I., Rabhi, F., & Motahari-Nezhad, H. R. (2022). Designing and managing human-AI interactions. *Information Systems Frontiers*, 24(3), 691-697.

[2] Weisz, J. D., Muller, M., He, J., & Houde, S. (2023). Toward General Design Principles for Generative AI Applications. *arXiv preprint arXiv:2301.05578*.

[3] Gmeiner, F., Holstein, K., & Martelaro, N. (2022). Team Learning as a Lens for Designing Human-AI Co-Creative Systems. *arXiv preprint arXiv:2207.02996*.

Conclusion

AI Algorithm Selection

AI Model	Pricing	Advantage
Stable Diffusion (SD)	Open Source	Free Allocation ★
Midijourney	Business	Visual Effect
Dall_E2	Business	Editable
Novel_AI (SD)	Business	Story Scene Generation

Table 2.5-1 Brief Comparison Between AI Models

The use of stable diffusion offers users increased freedom and flexibility. In light of this, the author chose stable diffusion as the basis for developing the prototype.

Human-AI Interaction

1. Use schematic framework for supporting co-creative human–AI team learning comprising core capabilities supporting creative processes and tool usage, and adaptive support layer for orchestrating creative collaboration between user and system by building shared mental models
2. Pay more attention to two of those challenges, human-AI conversations and collaboration, as well as explainability, accountability, fairness and bias
3. Gain valuable insights into how professional users learn to collaborate with existing generative AI tools with formative think-aloud studies

Reference (26)

Chapter 2

2.1 Reference

- [1] Weng, L. (2021, July 11). What are Diffusion Models? Lil'Log. <https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>
- [2] Lyu, S. (2019, June 26). Detecting deepfakes by looking closely reveals a way to protect against them. phys.org. <https://phys.org/news/2019-06-deepfakes-reveals.html>

2.3 Stable Diffusion Reference

- [1] Wikipedia contributors. (2023). Stable Diffusion. Wikipedia. https://en.wikipedia.org/wiki/Stable_Diffusion
- [2] Lewis, N. (2022b). How to Run Stable Diffusion on Your PC to Generate AI Images. How-To Geek. <https://www.howtogeek.com/830179/how-to-run-stable-diffusion-on-your-pc-to-generate-ai-images/>
- [3] Wuxiaojun. (n.d.). Diagram diffusion model: Understanding stable diffusion from the perspective of model. Zhihu column. <https://zhuanlan.zhihu.com/p/581186398>
- [4] Effective and efficient diffusion. (n.d.). https://huggingface.co/docs/diffusers/stable_diffusion
- [5] Lewis, N. (2022c). How to Run Stable Diffusion on Your PC to Generate AI Images. How-To Geek. <https://www.howtogeek.com/830179/how-to-run-stable-diffusion-on-your-pc-to-generate-ai-images/>
- [6] Ma, Y. (2023). The 10 Best Stable Diffusion Models by Popularity (SD Models Explained). AiTuts. <https://aituts.com/models/>
- [7] Boehman, C. (2023). Midjourney vs. DALL-E vs. Stable Diffusion: Which Is Better? MUO. <https://www.makeuseof.com/midjourney-vs-dalle-vs-stable-diffusion/>

2.3 DALL_E2 Reference

- [1] Wikipedia contributors. (2023a). DALL-E. Wikipedia. <https://en.wikipedia.org/wiki/DALL-E>
- [2] Kerner, S. M. (2023). Dall-E. Enterprise AI. <https://www.techtarget.com/searchenterpriseai/definition/Dall-E>
- [3] What is DALL-E 2? Use Cases and Examples. (2023, January 2). Cubix. <https://www.cubix.co/blog/what-is-dall-e-2>

2.3 Midjourney Reference

- [1] Wikipedia contributors. (2023a). DALL-E. Wikipedia. <https://en.wikipedia.org/wiki/DALL-E>
- [2] Hughes, A. (2023). Midjourney: The gothic AI image generator challenging the art industry. www.sciencefocus.com. <https://www.sciencefocus.com/future-technology/midjourney/>
- [3] Urrutia, D. (2023). What is Midjourney | Definition, advantages and how to use it. Arimetrics. <https://www.arimetrics.com/en/digital-glossary/midjourney>
- [4] Boehman, C. (2023b). Midjourney vs. DALL-E vs. Stable Diffusion: Which Is Better? MUO. <https://www.makeuseof.com/midjourney-vs-dalle-vs-stable-diffusion/>
- [5] Melekhin, K. (2023, March 15). What is Midjourney? A Comprehensive Guide. HostAdvice. <https://hk.hostadvice.com/blog/ai/what-is-midjourney/>

2.3 Novel_AI Reference

- [1] Goodwin, R. (2023, June 6). What is Novel AI? Features & How It Works Detailed. Know Your Mobile. <https://www.knowyourmobile.com/ai/novel-ai-features/>
- [2] NovelAI - The AI Storyteller. (n.d.). NovelAI. <https://novelai.net/>
- [3] Visimens-Technology. (2022, October 14). What is a novel ai and how to use it - visimens technology. visimens technology. <https://visimens-technology.com/what-is-a-novel-ai-and-how-to-use-it/>
- [4] Paiva, R. (2023, March 23). What is NovelAI? Review - Robot Artificial. Robot Artificial. <https://www.robotartificial.com/what-is-novelai-review/>
- [5] The_Human_Empire. (n.d.). Reddit - Dive into anything. https://www.reddit.com/r/NovelAi/comments/oe16kf/what_exactly_is_novelai/
- [6] Chainanon. (2023). NovelAI Review: How good is the AI storyteller really? AiTuts. <https://aituts.com/ultimate-novel-ai-review/>

2.4 Reference

- [1] Abedin, B., Meske, C., Junglas, I., Rabhi, F., & Motahari-Nezhad, H. R. (2022). Designing and managing human-AI interactions. Information Systems Frontiers, 24(3), 691-697.
- [2] Weisz, J. D., Muller, M., He, J., & Houde, S. (2023). Toward General Design Principles for Generative AI Applications. arXiv preprint arXiv:2301.05578.
- [3] Gmeiner, F., Holstein, K., & Martelaro, N. (2022). Team Learning as a Lens for Designing Human-AI Co-Creative Systems. arXiv preprint arXiv:2207.02996.

03

Mannequin Tool Research



Chapter 3 presents the product analysis and user research insights on the Mannequin Tool (MT), which shows an overview of the research field and helps define the project focus.

Chapter Overview

3.1 Product Build-up

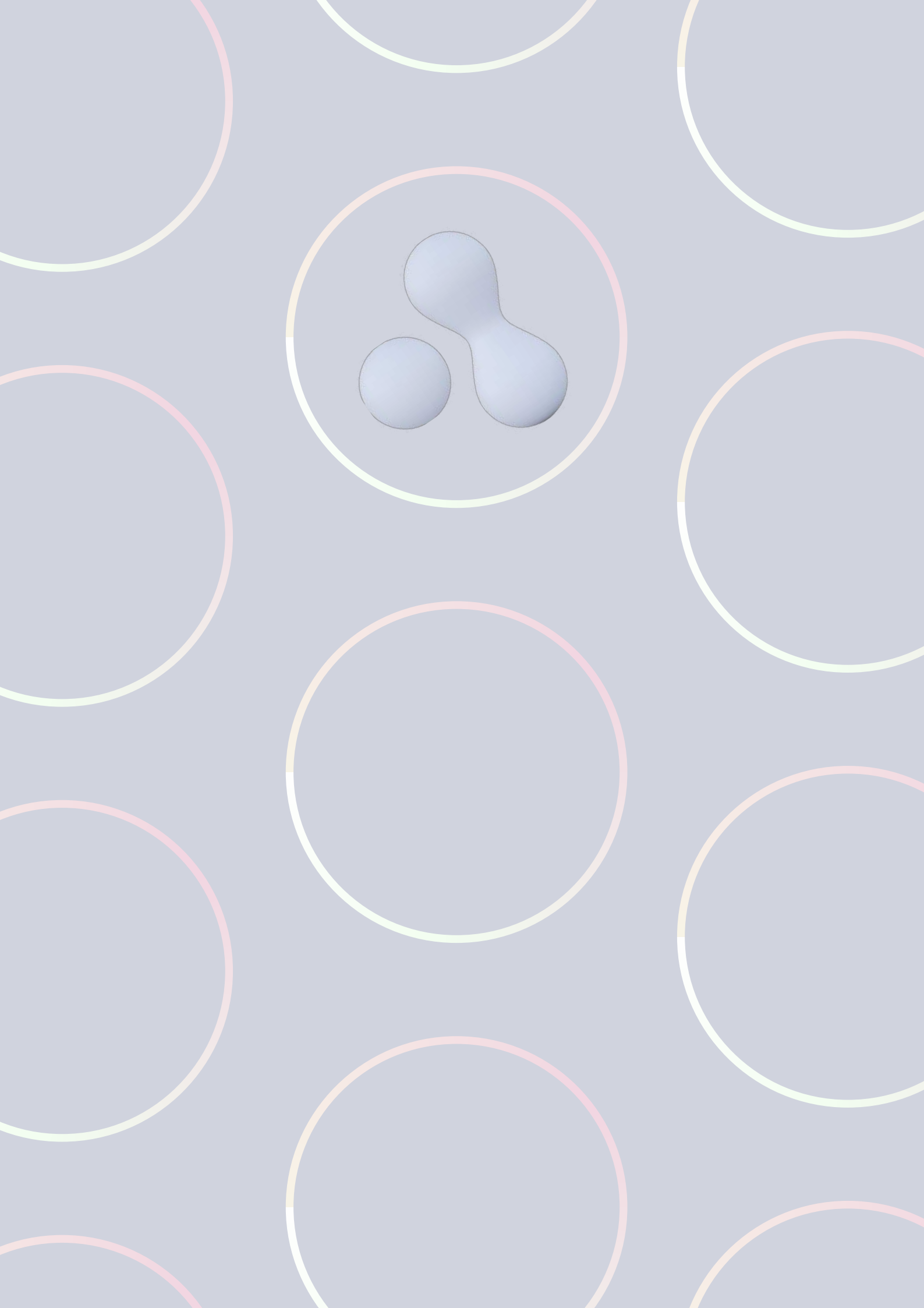
3.2 Explorative Interviews

3.3 Interview Data Analysis

3.4 Context Mapping & Analyzing The Data

3.5 Thematic Analysis Results

Key Insights



Chapter 3

3.1 Reference

[1] Chou, J. (2022, December 16). Flow Chart - Deerlight Design - Medium. Medium.<https://medium.com/deerlight/design-755bfb988b9>

3.1 Product Build-up



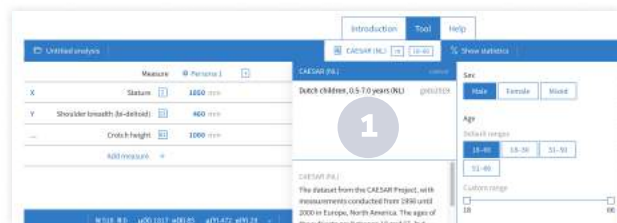
Figure 3.1-1 Official website explanation

There are two populations that can be selected.

CAESAR



Dutch children



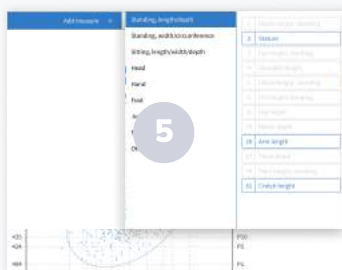
Name



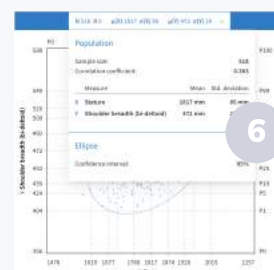
Show Statistics



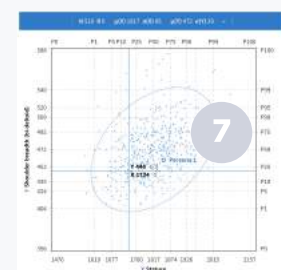
Download



Add measurements



Input Information



Interactive Scatter Plot

Figure 3.1-2 Main Functions and UI

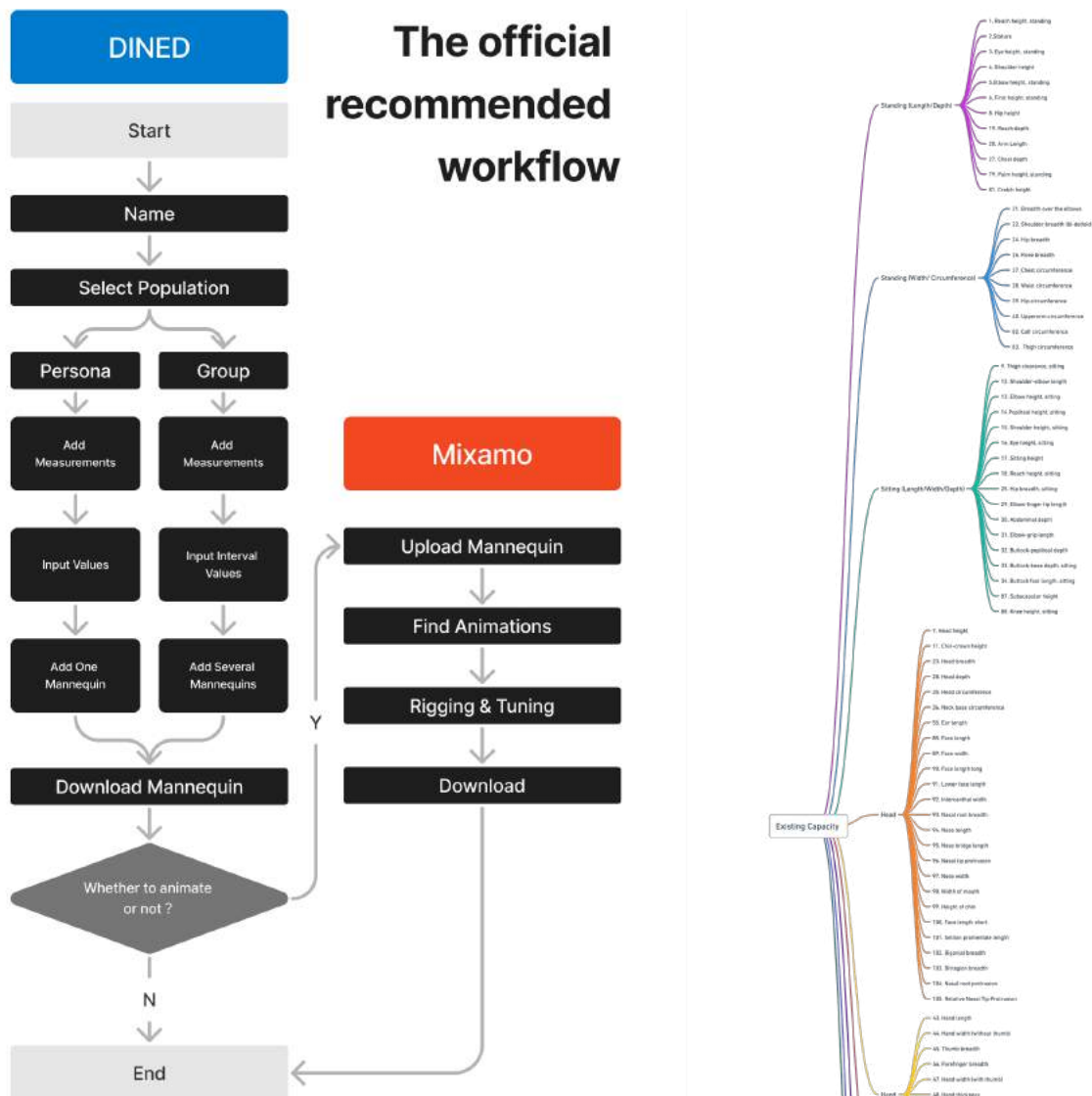


Figure 3.1-3 The Official Recommended Workflow [1]

The DINED/Mannequin Tool user interface is provided as a website. It consists of a population selection and specification panel, a measurement selection and manikin creation panel, a scatter plotting panel for exploring the data distribution and to view the position of the manikins in the population, a panel for interactive 3D viewing of the manikins, and a module to download manikin data, scatter plots, and STL-files of the manikins. <https://www.dined.nl/en/how-it-works/3d#mannequin> Manikins generated with DINED Mannequin can also be animated, e.g. for use in presentations or for virtual testing of products. On the right, a P50 male manikin was animated using Mixamo



Chapter 3

3.2 Reference

- [1] A. M. Hickey and A. M. Davis, "Elicitation technique selection: how do experts do it?," Proceedings. 11th IEEE International Requirements Engineering Conference, 2003., Monterey Bay, CA, USA, 2003, pp. 169-178, doi: 10.1109/ICRE.2003.1232748.
- [2] Frow, P., Nenonen, S., Payne, A., & Storbacka, K. (2015). Managing co-creation design: A strategic approach to innovation. British journal of management, 26(3), 463-483.

3.2 Explorative Interviews

This session describes the empirical study conducted by the author, which includes the background, used methods, procedure, and findings. (Please See Appendix C3.2)

Background:

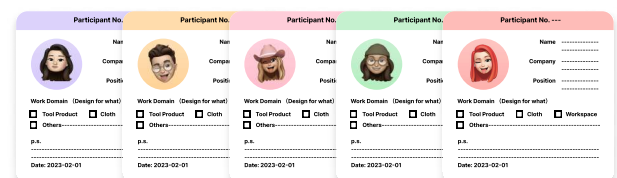
The previous research has unveiled the main theoretical foundations that informed this project and illustrated the understanding of the state-of-the-art (SOTA) of human-form AI algorithms. While the promising AI algorithms and the development direction were identified in the literature research, a deeper understanding of users from a pragmatic perspective was needed. Therefore, five semi-structured interviews with the DINED/Mannequin tool users were conducted to collect feedback and insights about their experiences of using the DINED/Mannequin tool or doing an ergonomic project and attempting to find potential opportunities for emerging human form generative AI models soon.

Method :

The user interview process consists of two parts: the first part is a story elicitation interview[1] where the interviewer asks the interviewee about their experience with the Mannequin Tool and how they use it in their workflow. After the warm-up session, participants are encouraged to share their last tool usage story. The second part is a co-creation session (Pennie Frow et al., 2015) [2] where the interviewer introduces the interviewee to some emerging 3D human form-related AI models based on the previous AI literature review and asks them which one they would like to try in their workflow and how they would integrate it into their process.

The main aims of the online one-on-one interview were:

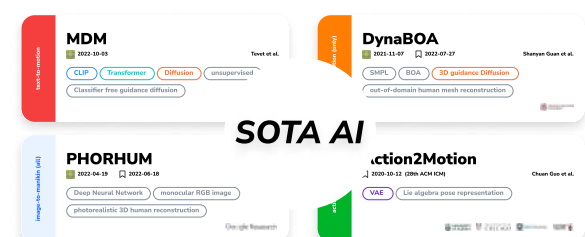
1. To learn the richness of their work experience in anthropometry;



2. To understand how they involve the DINED/Mannequin tool or some other anthropometric/ergonomic tools in their workflow;



3. To intrigue participants with SOTA AI in co-creation session, brainstorm design ideas for the DINED/Mannequin tool;



Participant Screening and Recruitment

The screening criteria were established on the DINED/Mannequin Tool usage experience and the demand for anthropometric/ergonomic tools/methods. This user interview paid more attention to MT users who have a strong need to use the ergonomic product but do not often use the DINED/Mannequin Tool. The recruitment process was done by utilizing the networks of Dr. Toon Huysmans and the author.

No.	Name	Occupancy	Background
1	Miss. A	Ms. Student & Quasi-Ph.D.	She has determined her future Ph.D. research direction to be comfort ergonomics.
2	Miss. T	Intern & Ms. Student	She is doing an internship at a Danish company that produces medical consumables intended for direct application to the human body.
3	Mr. T	Wetsuit & Softgoods Developer	He used doing research for the founding of the DINED/ Mannequin tool.
4	Miss. L	Scientist Innovator	She does research concerning anthropometry for the military.
5	Miss. S	Ex-researcher in Philips	She researched on breast pump in Philips and holds an interest in the anthropometry of women during lactation.

Table 3.2-1 Demographic of Participants

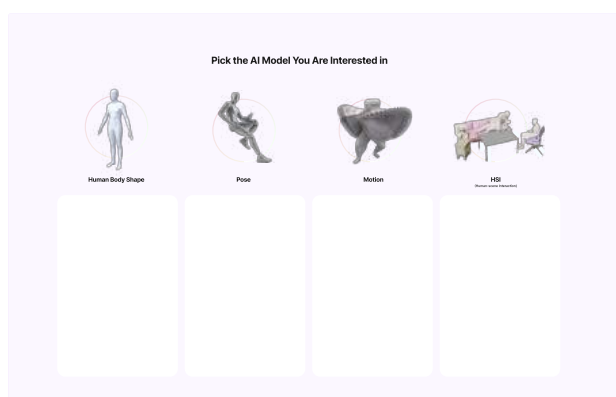


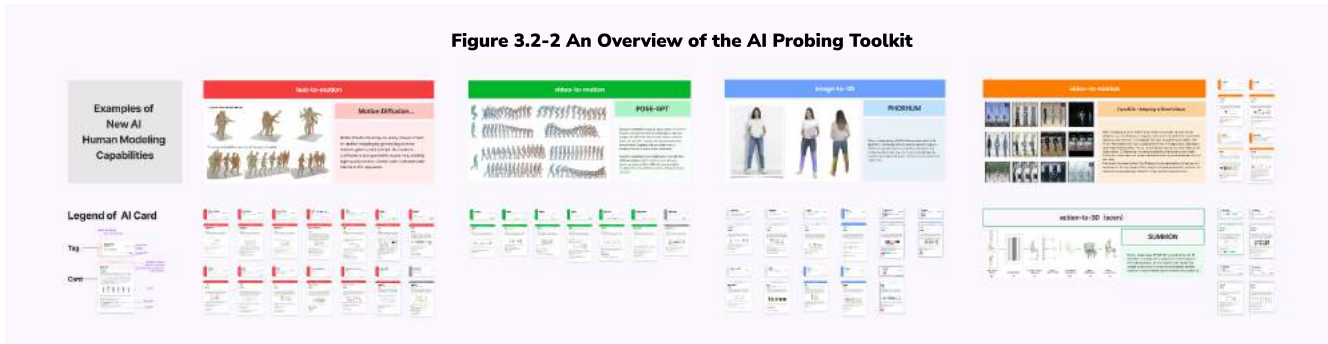
Figure 3.2-1 Guiding Directions: Human body, Pose, Motion, HSI

Data Collection Guide

Combining AI capabilities with the usage purpose of the DINED/Mannequin tool, the interviewer proposed four distinct aspects of generating

objects: body shape, pose, motion, and HSI (Human-scene Interaction). As the conversation progressed, participants engaged in a meaningful dialogue about their usage requirements, existing product defects, and expected usage effects of the tool. This enabled a more comprehensive understanding of how to best utilize the DINED/Mannequin tool and how to co-create AI products that would be able to meet their needs. Moreover, participants were able to identify potential areas for improvement, and discuss future plans for the development and utilization of the tool. This discussion provided a valuable opportunity for everyone to contribute to the development of DINED/Mannequin tool, and to engage in the co-creation of AI products.

Figure 3.2-2 An Overview of the AI Probing Toolkit



Data Collection AI Probing Toolkit

A set of cards (Appendix X) were designed to serve as a co-creation starter. The cards are categorized according to the capabilities of the AI algorithms which are researched in the literature review. It could help the participants better think of the potential anthropometry functions. While leading participants to learn the SOTA AI algorithms, the interviewer encouraged participants to think aloud. And inquiry questions were asked to gently guide the conversation and reveal the desires and needs.

Procedure:

All interviews were conducted online considering that the interviewer and all the interviewees were working in different cities and the author is better at designing online whiteboards to assist interviews. In each 45-minute interview, the interviewer asked participants questions about their professional experience. Table 2.3-2 shows the overview of the interview.

Opening:

The participants were asked to read and sign the consent form. After that, the interviewer briefly introduced the background of the project and let the participant ask questions before officially starting. With their permission, the interviewer recorded participants on audio and/or video.

Section One:

Warm-up

Each participant was first invited to talk about what

he or she is busy with now and how they know the tool DINED/Mannequin.

User Story

The participant was encouraged to remember the last time him or her used the anthropometry/ergonomic tool/method. If the interviewee found it difficult to recall a story then let he or she tell about a time when anthropometry 2D or 3D data is needed in the work.

Share Some Concrete Uses

The participant was asked to share a screenshot of the usage of the tool, if applicable. And then, centered on the concrete uses and experience story, the prepared questions were elicited to dig for more information.

Section Two: Co-creation1

Co-creation is a collaborative initiative between companies and their customers enabling the joint design of products and services. These initiatives include the creation of goods, services and experiences, amplifying the process via the inclusion of client intellectual capital. (<https://www.gartner.com/en/information-technology/glossary/co-creation>)

The participant was shown four states of the 3D human body, in which a series of SOTA AI algorithms related to different 3D human forms were listed. The participant was asked to associate one or several functions with the AI algorithm that was most desired and to be endowed with extra functions. Participants could choose more than one AI algorithm and share their thoughts on how they would like to use the AI ability in what kind of context.

Aim		Activity		
To learn the work experience in anthropometry	How users use the tool?	Warm-up & Ice-breaking	Usage Story	Interface Screenshot
	What can AI help with?	Co-Creation	Pick AI Cards	Think Aloud
SECTION ONE	Greetings	Introduce / Sign Consent Form / Warm-up Questions (7min)	-Ask for permission to record -Do you have any questions before we start?	
	To learn how users use the tool? (Interviewee lead)	User Story (4min) Show Some Concrete Uses (3min) Guiding Questions (20min)	-Share an experience about the last time you used the anthropometry 3D digital model generated by Mannequin Tool? -Could you please share your screen and show me how you generate the 3D digital model? (screenshot)	
	To learn how users use the tool? (Interviewer lead with guiding questions)		1. What is that 3D digital model used for? 2.In your workflow when will you use the Mannequin Tool? 3.How you integrate the tool in your design/research process? 4. What activities will you do before opening the DINED tool? 5. How will you deal with the outcome you get from the Mannequin Tool (For example, animating the manikins in Mixamo or importing model in CAD)? 6. What do you think is the key outcome in your expectation of the Mannequin Tool? Why is that important to you? 7. What other ergonomic tools do you use? And how do those tools cooperate with the Mannequin Tool? 8. Do you need to prepare something before you use the Mannequin Tool? What do you need to prepare? 9. Have you faced any problems while working with the Mannequin Tool in this process or before? / Have you ever faced any problems while working with the Mannequin Tool ? What kind of solutions did you try? / Or what's prevented you from looking for a solution?	
SECTION TWO	To understand what can AI help with?	Co-creation: human form generative AI models (10min)	In the current context of accelerated integration of the digital and physical worlds, generative AI models are quietly leading a profound change, reshaping or even overturning the production and interaction modes of digital content. Here are some emerging 3D human form related AI models. Whether there is value for your work ? According to your own using experience, which AI models would you like to try in your work flow? And how you would integrate it in your process. 1. Read AI Reference (see the picture of section 2: Co-creation), 2. Pick the AI model you are interested in maybe also tell us why you are interested in this model?	
	Thank you very much for offering your user experience and opinions !			

Table 3.2-2 An Overview of the Interview Process

3.3 Interview Data Analysis

Brief User Individual Contexts

I interviewed five DINED users and their experiences were varied, yielding many insights. I present their user profiles in terms of usage frequency and related demand. Target groups can be generated from the analysis of the current situation of Mr.T, Miss. A, and Miss. T, who use the tool more often and whose requirements for anthropometric/ergonomic tools/methods do not exceed the DINED/Mannequin tool's capabilities too much. Since this tool already meets the needs of this user well in terms of function. The trend of user needs can be seen in the significant needs shared among Miss.S, Miss. L, Miss. A, and Miss. T.

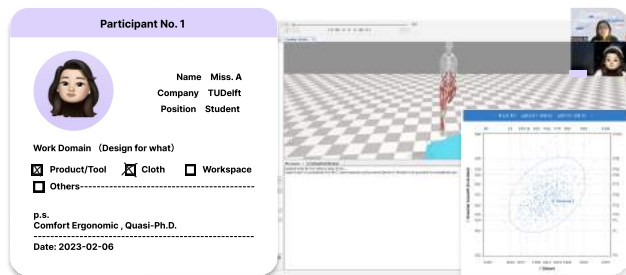


Figure 3.3-1 Interview with Miss. A

Context of Miss. A:

She focuses on researching comfort ergonomic design and has a great need for ergonomic tools. She uses the Mannequin Tool for Ellipse chart. As for testing eyeglass frame design and observing stress on different body parts, however, she only uses it a few times and typically relies on other products. She faces some limitations with the Mannequin Tool, particularly with the adjustment options during the eyeglass frame design process, so she opts to use other software tool, Paraview,

instead of trying to find solutions with the Mannequin Tool. It is emphasized that there is a need for data applicability.



Figure 3.3-2 Interview with Miss. T

Context of Miss. T:

She is a Master's student majoring in Integrated Product Design. She is doing an internship with a Danish company that produces medical consumables intended for direct application to the human body. She used mannequins of various fitting movements to design the fitting room space. Though she just used the DINED/Mannequin tool in the course, she is able to reflect several ergonomic user needs. She suggests using technology like 3D scanning and modeling software to simulate the body profiles and scars of patients, which can help in designing medical products. She also proposes using ergonomic plug-ins in modeling software to save time during the iteration process. Additionally, she suggests simulating different body weights and contact areas to test the durability of products and using virtual or augmented reality in the design process. Finally, she mentions the potential for software to explain complicated subjects like biometric mechanics and aid in the calculation process, as well as using pose and motion data to design sports equipment.

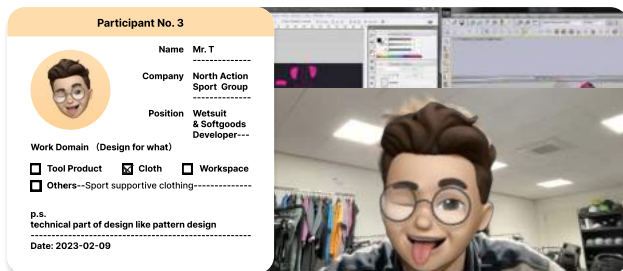


Figure 3.3-3 Interview with Mr. T

Context of Mr. T:

Mr. T, a wetsuit and softgoods developer, is the most ideal DINED/Mannequin user. He uses the DINED/Mannequin tool in each work process and does not need to use any other anthropometric tool or ergonomic method. He mentioned that their company's design process starts with a medium-sized 3D human digital model as the base, and then they make adjustments to create other sizes, including large, tall, large, and small. They use physical samples to test the fit of the wet suits, but they also use digital testing methods such as creating a 3D pattern and using a virtual model to simulate how the wet suit will fit. The goal of this process is to make the final product fit perfectly, and it involves multiple rounds of updates and adjustments until the person is satisfied with the design.

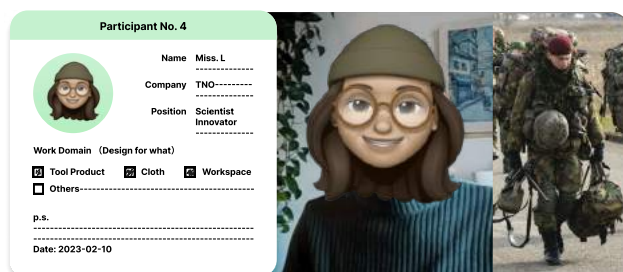


Figure 3.3-4 Interview with Miss. L

Context of Miss. L:

She is a scientist innovator in the military. She will not use any external software under military confidentiality requirements. She highlights the importance of taking into consideration the

clothing and equipment worn by individuals in research. She mentions that just taking into account the body measurements is not enough, as it does not account for the added thickness and weight from clothing and equipment. The speaker mentions a specific example of a bulletproof vest, which can affect the way someone sits in a chair. She suggests that there is a gap in knowledge in this area and would like to have a more integrated approach where a digital model of a human can be dressed with different equipment to see the impact on posture and sitting angle. This would allow for a more accurate representation of how individuals would sit in real-life scenarios.



Figure 3.3-5 Interview with Miss. S

Context of Miss. S:

She just finished a project that cooperated with Philips company that was focused on testing the sensitivity of the chest in the lactation period using a biomechanical model made of silicone material. She and her colleague iterated the project several times and used a self-built database to calculate the approximate distribution data. She adopted the method from a previous Ph.D. project that was focused on diagnosing chest tumors and imitated it to make a prosthesis. They used silicone to make the skin surface. The silicone model was used to test the feasibility of the scanning method, test the pressure on the chest under different conditions, and demonstrate the chest sensitivity data to participants.

3.4 Reference

[1] Visser, F. S., Stappers, P. J., Van der Lugt, R., & Sanders, E. B. (2005). Contextmapping: experiences from practice. *CoDesign*, 1(2), 119-149.

3.4 Context mapping and Analyzing the Data

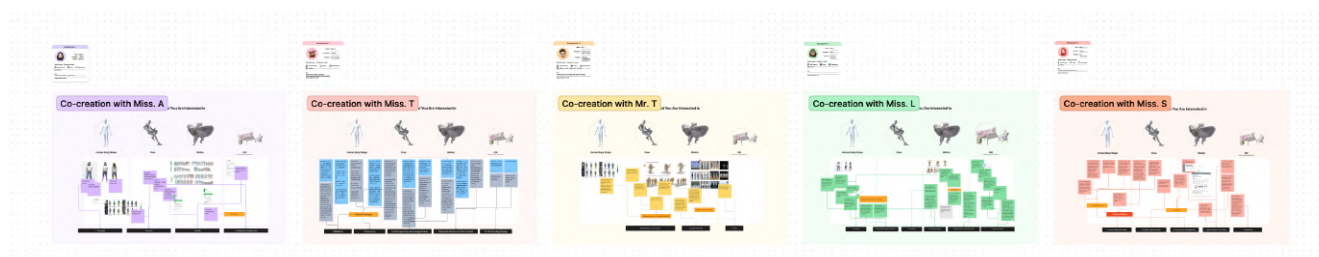


Figure 3.4-1 Initial Quote Mapping (Please See Appendix C3.4)

The recording of the interviews was first transcribed and then the researcher first obtained the initial insights. The transition from the data to themes was completed by **contextmapping** [1], which is clustering the initial quotes and analyzing them with extended themes that categorize the raw data. The analysis methods used here are thematic analysis, frequency, and discourse analysis. The identified **twelve themes** clustered into **three groups** were listed below.

Group 1: Product Quality Requirements

Product quality requirements refer to the specifications, features, and characteristics that DINED/Mannequin tool should possess to meet customers' needs and expectations. (T=Theme)

T-2: Accuracy This theme refers to user needs about how closely a 3D human body model matches the real-world object it is meant to represent. In other words, it cares about the real dataset and how well the 3D model captures the shape, dimensions, and other physical attributes of the anthropometry.

T-4.1: Easy This theme reflects the user's need of simplifying the 3D model generation method.

T-6: Quality of 3D Model This theme focuses on the quality of the model, which is, less broken mesh faces 3D model.

Group 2: Product Design Development Directions

Product design development directions refer to the various approaches or strategies that can be used in the improvement of DINED/Mannequin.

T-1: Personalized This theme leads to personalized design which is a design approach that focuses on tailoring products, services, or experiences to the specific needs and preferences of individual users or customers;

T-3: Editable This theme is about allowing the generated model to be manually edited to meet the needs of more scenarios;

T-4: Knowledge Guide Offer guidance to users about how to use the anthropometry data. where to find the anthropometry data and so on;

T-4.2: Dataset (Missing Data) Offer solution when users cannot find the measurements in DINED;

T-5: Specific Motions in the Context The typical motions that are expected or required in a particular context may vary based on cultural, social, or environmental factors.

T-7: Equipment In some contexts, besides human body data, the equipment is important for measurements.

T-9: Simulation Software Function This theme refers to the user's need of creating virtual models of real-world systems, allowing users to test and analyze various scenarios in a safe and controlled environment.

T-12: VR Assisting Design Virtual Reality (VR) can be a powerful tool in assisting ergonomic design processes.

p.s.: Both **Easy** and **Dataset (Missing Data)** can be solved by setting a proper anthropometry knowledge guide. Thus, those two themes are sorted in **T4-Anthropometry Knowledge Guide**.

Group 3: Considerations

T-8: Security Problem A security problem is any issue or vulnerability that can potentially compromise the security of a system, network, device, or data.

T-10: Human Body Ethical Problem The human body raises many ethical concerns, ranging from using human subjects in medical research to end-of-life decisions. Here are a few examples of ethical issues related to the human body.

T-11: Female Care This theme refers to the care for health and hygiene products, services, and practices that are specifically designed for or marketed toward women.

Please See Appendix C3.4

Clustering and Analyzing the Interview Data

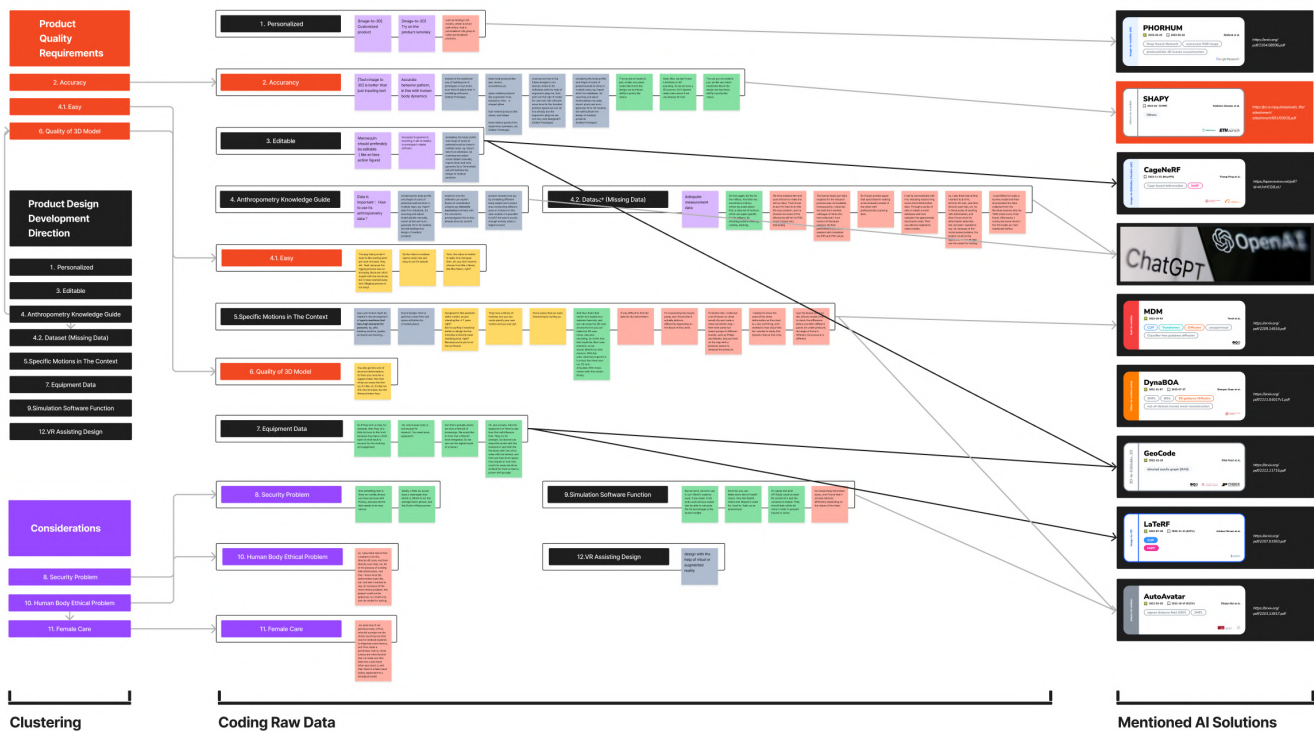


Figure 3.4-1 Initial Quote Mapping <https://www.figma.com/file/hwKEjINaeloooEJldAawCg/Clustering?node-id=0%3A1&t=V83WxAB3Mneu2JJS-1>

3.5 Thematic Analysis Results

Findings from Group 1

Product Quality Requirements

Based on the interviews conducted, it was discovered that users place great importance on the **accuracy** of human body data and the **quality** of 3D or motion representations. Therefore, when utilizing AI generation capabilities, it is essential to consider how to precisely utilize the data to accurately restore the human body, pose, and motion. Additionally, users anticipate that the generation capabilities of the tool will be user-friendly, **easy**, and efficient.

“ [Text+image to 3D] is better than just inputting vague and uncertain text to achieve good effects. ”

T2-Accuracy, From Miss. A

“ The rigging process was so annoying. Because I did it myself with the Mixamo, but it never worked really well. ”

T4.1-Easy, From Mr. T

“ You also got like a lot of incorrect deformations. So then you have like a rigged model. But then when you move the Arm up, it's like, uh, it's like not the natural shape, but the Always broken face. ”

T6-Quality of 3D, From Mr. T

Findings from Group 2

Product Design Development Directions

According to the frequency of quotes in the second cluster, the most frequently mentioned development directions are **Anthropometry Knowledge Guide, Specific Motions in the Context, and Equipment Data**.

In the context of the rapid development of AI, a good anthropometry tool could and should be able to inform users with expertise in many ways.

Such a tool should be able to provide guides on the expected 3D/4D human form outcomes.

“ It would be nice if the software can explain the basics of complicated subjects eg. biometric mechanics. ”

T4-Anthropometry Knowledge Guide, From Miss. T

“ Data is important but sometimes I do not know how to use the data in DINED. ”

T4-Anthropometry Knowledge Guide, From Miss. A

“ The human body part data required for the research process was not available. Consequently, I obtained the data from another colleague of mine who had collected it from dozens of literature sources. He then performed the statistical analysis and calculated the 95% and 50% values. ”

T4.2-Dataset (Missing Data), From Miss. S

When working on ergonomic projects, users often require dynamic data for tasks such as deformation and motion. Unfortunately, there is currently no both reliable and convenient solution available to meet these needs. Without a deep understanding of complex professional simulation software, users must rely on their own experience and estimation to solve problems. Learning anthropometry can be a daunting task for beginners in the field, who may find it challenging to acquire and master the necessary skills without proper guidance. This can have a significant impact on the tool's reusability for beginners in their subsequent learning endeavors. Without the support of a professional instructor, users may avoid engaging in rigorous discussions on this aspect and lack the motivation to use the tool actively. Additionally, the complexity of anthropometry design is compounded by the fact that **different contexts require different types of motion**, further adding to the difficulty of the process.

“ It is not easy to design with aesthetic static motion, so just a standing T pose, right? But for surfing it would be better to design with a little lower standing pose, right? ”

T5-Specific Motions in the Context, From Mr. T

“ For the military, we would like to have a library where we could select from a data set of motions that are super specific for the military such as shooting posters crawling, running, and ducking. ”

T5-Specific Motions in the Context, From Miss. L

In the research process, relying solely on human data may not always suffice to meet research needs. Factors such as the designed product, the everyday equipment used by the target group, and the operating environment can also play a significant role in the research outcomes. To address this, users seek a straightforward solution that combines human data with equipment and environmental data, in order to obtain a more complete picture of the research subject.

“ So if they sit in a chair, for example, then they sit a little bit more to the front because they have a thick layer on their back to account for the clothing and equipment. ”

T7-Equipment Data, From Miss. L

According to the feedback of participants, editable 3D is a promising direction that should be considered in any idea. The author emphasizes the importance of focusing on designing for interaction, rather than solely on technological realization. Personalized design is an approach commonly used in the anthropometric design. This methodology can serve as a benchmark for assessing the usability of an idea during subsequent prototyping stages. While the development of simulation software and VR-assisted design functions for the DINED tool is a challenging task that falls beyond the scope of this project, the author will address strategic planning related to these two directions in future work discussions.

Findings from Group 3

Considerations

When incorporating AI technology into tools, it is crucial for developers to prioritize the protection of user data privacy and to be mindful of potential ethical issues related to the human body. For instance, it is important to avoid making derogatory comments about body shapes and to refrain from triggering body image insecurities.

“ And something else is there on mobile phones you have an issue with privacy because all the data needs to be very secure. ”

T8-Security Problem, From Miss. L

“ There is very little data overall about women in these aspects. Only breast enlargement surgery papers will include. ”

T11-Female Care, From Miss. S

Workflow Inspiration

Interviews revealed that the workflow for 3D human body data models can be summarized into four parts, regardless of the context: researching target group, researching data or a database, inputting the data, creating 3D/4D models, and refining the model.

Users usually use other software to supplement the model editing part. Given that the development stage of the DINED product is in the initial stage, the design of complex and diverse simulation functions can be considered in the longer-term planning and is not within the scope of consideration of the current project.



Figure 3.5-1 Overview of Workflow

Key Insights

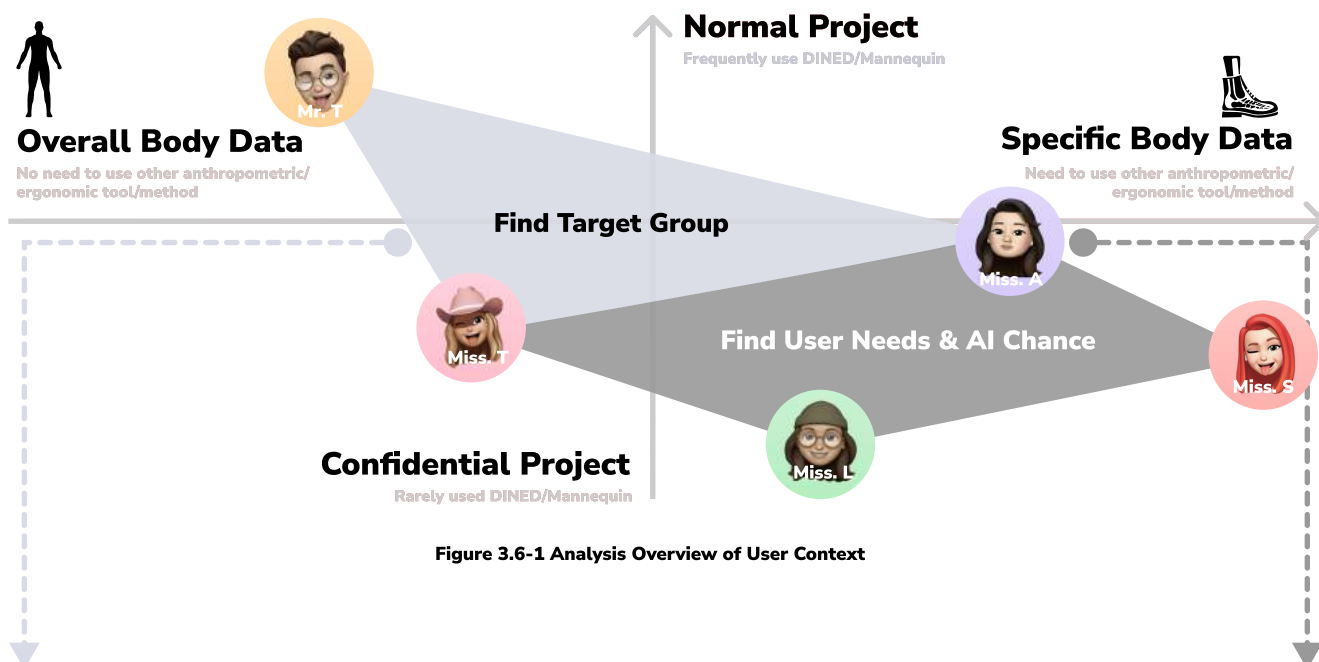


Figure 3.6-1 Analysis Overview of User Context

Target Group



Users concerning diverse body shapes

Users involved in general human factor projects related to the **overall body data** of the diverse body shapes, rather than specific parts of the body or confidential projects.

TOP 3 Function Requirements



Knowledge Guide

Offer guidance to tool users about how to use the anthropometry data, where to find the anthropometry data and so on;



Specific Motions in the Context

The typical motions that are expected or required in a particular context may vary based on cultural, social, or environmental factors.



Equipment

In some contexts, besides human body data, the equipment is important for measurements.

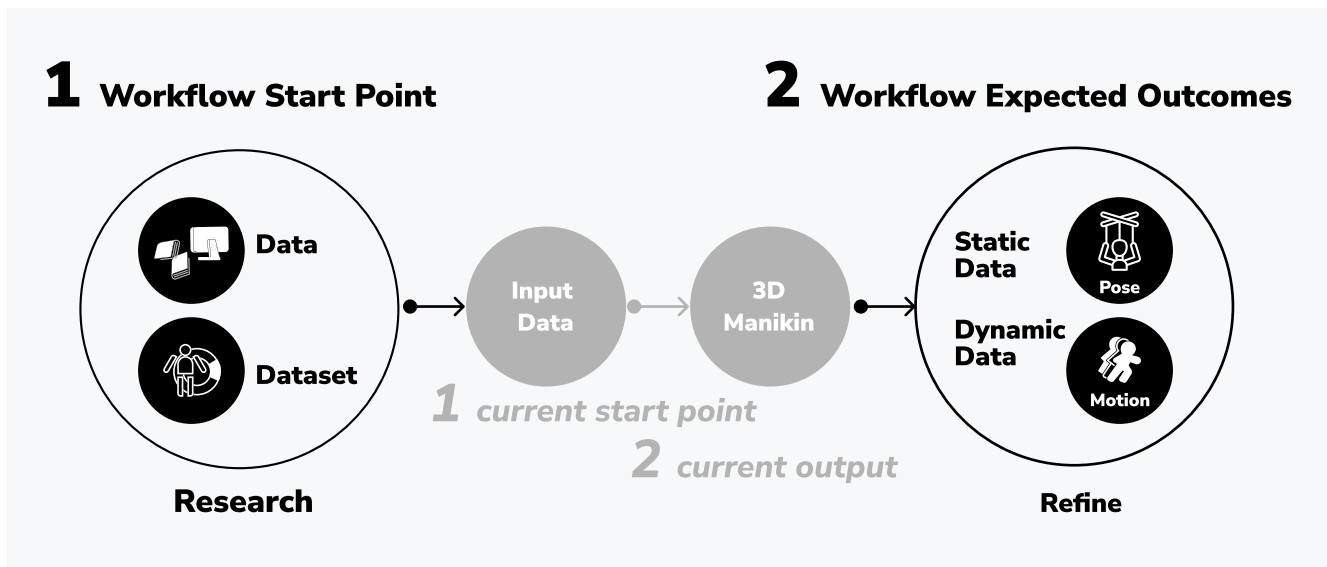


Figure 3.6-2 Workflow Optimization

Workflow Based Optimization

Based on our analysis of user workflows, it is discovered that a significant level of similarity in the processes involved in using DINED products and competitive products. Specifically, before utilizing any data, users commonly conduct thorough research on human body data and anthropometry datasets. After generating a 3D model, rigging and human motion generation are highly demanded by users.

As such, our project's primary focus for design research is to enhance the workflow and make it even more user-friendly. This will involve providing novice users with valuable anthropological knowledge during the early research stages. Additionally, we will address the high user demand for functions such as rigging and motion generation for 3D manikins.

Requirements for the overall general experience of the product

When utilizing AI generation capabilities:

1. It is essential to consider how to input the data to accurately restore the human body, pose, and motion and ensure output target measurements precisely;
2. It is anticipated by users that the generation operations of the tool should be **easy**;
3. It is crucial to prioritize the protection of user data privacy and to be mindful of potential ethical issues related to the human body;

Product Quality Requirements

T-2: Accuracy

T-4.1: Easy

T-6: Quality of 3D Model

T-8: Security Problem

T-10: Human Body Ethical Problem

Product Quality Requirements

04

Conceptualization



Chapter 4 follows the literature review and field study, where I conducted an intensive brainstorming session to generate a multitude of concept ideas. Subsequently, I developed an example case and defined the specific design goal to be pursued.

Chapter Overview

4.1 Brainstorming Workshop

4.2 Brainstorming Insights

4.3 Insights Combined With User Research

4.4 Final Concept

4.5 Example Case

4.6 Design Goal

Reference

4.1 Brainstorming Workshop

A co-creation workshop was held to generate initial DINED AI ideas that could potentially achieve the intended achievable effects in design process. In other words, this brainstorming workshop is to figure out how AI can help with **diverse body shape related design** combined with the abilities of DINED. Also, the author wants to learn the current and potential AI application in design market.

Background:

Through an AI communication interest group, the author learned in advance before the workshop that "stable diffusion" and "midijourney" have relatively high popularity in the design application. Many game companies and internet companies are researching how to efficiently produce high-quality original artwork materials and graphic design materials through AI. Therefore, in this workshop, the author uses "stable diffusion" as an example to attract the interest of designers and stimulate their inspiration.

Method :

In this workshop, the "Simulation Experience [1] + Divergent Thinking [2]" model was adopted. Due to the free open-source ControlNet algorithm of Stable Diffusion (SD), which brings more possibilities to SD, the author selected SD as the product to be experienced and deployed SD with ControlNet locally on her personal computer. The participants are allowed to send input contents to the author online, who would then reply with content generated by SD in formative think-aloud studies (mentioned in Chapter 2.4). Through this artificial simulation of AI product, participants were able to experience the **prototype** and focus more on idea generation. As the process of AIGC product generation and artificial feedback is time-consuming, the author conducted two workshops (one with 4 participants and another with 3) and both brainstorming workshop last 2 hours.

Task Description:

1. Find Four pictures with human pose (two for workflow 1, two for workflow 2) (Suggestions : Avoid images with exaggerated pose or in small size, and the background should not be too complex)
2. Writing Prompts (Suggestions : [Stable Diffusion prompt: a definitive guide](#) - [Stable Diffusion Art](#) ([stable-diffusion-art.com](#)) [3])

Workflow 1: Generate an image with a similar human pose



Workflow 2: Image generation is controlled by body pose and stick figure



Figure 4.1-1 Workshop Plan (Appendix-4)

Participant Screening and Recruitment

The screening criteria were established on design background. This user interview paid more attention to designers who have got in touch with AI design market and at least had one year design work experience. The recruitment process was done by the author.

No.	Name	Occupancy	Attitude Towards Text-to-Image
1	Mr. D	Big Bank App Product Manager	"Can be used as product operation chart"
2	Mr. L	Mobile Phone Company Interaction Designer	"The company has carried out AI lectures, but our department has not applied them yet."
3	Miss. X	Game Company Interaction Designer	"An interaction designer with AI is powerful than visual designer ! "
4	Miss. Y	Game Company Technical Artist	"Mature Icon Generation"
5	Miss. C	IoT Company Digital Twin Product Manager	"So fun to learn AIGC and there are a lot insights can be found"
6	Miss. J	Internet Company AI Product Designer	"It seems that AI is more good at generate inspiration."
7	Miss. L	Amusement Park Company Designer	"The more I interact with the product, the more inspired I get with the personal setting drawings."

Table 4.1-1 Demographic of Participants

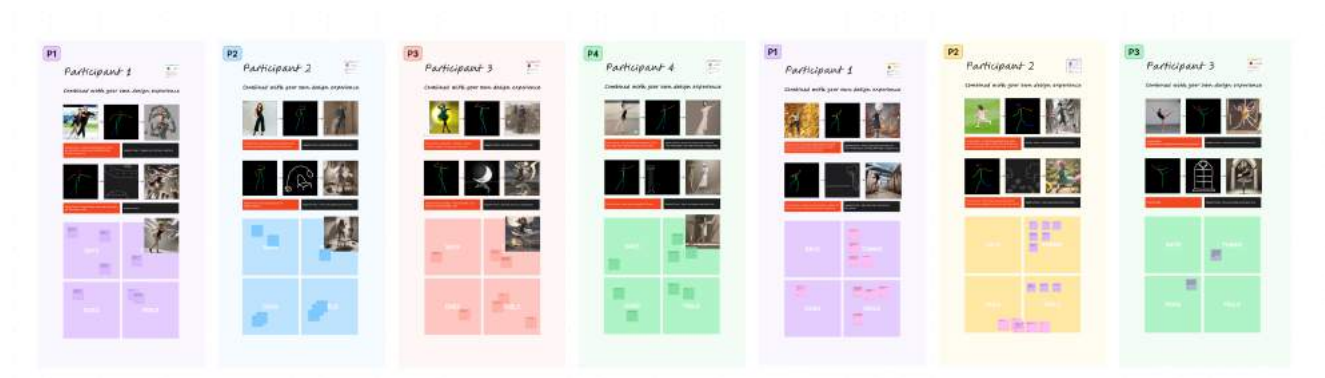


Figure 4.1-2 Online White Boards for Participants (Please See Appendix C4.1)

4.1 Reference

- [1] Müller, M., & Pfahl, D. (2008). Simulation methods. Guide to advanced empirical software engineering, 117-152.
- [2] Lee, J. H., & Ostwald, M. J. (2022). The relationship between divergent thinking and ideation in the conceptual design process. Design Studies, 79, 101089.
- [3] Andrew. (2023, May 15). Stable Diffusion prompt: a definitive guide - Stable Diffusion Art. Stable Diffusion Art. <https://stable-diffusion-art.com/prompt-guide/>

4.2 Brainstorming Insights

Feedback of the AI Generated Results

Here are the informing feedback and evaluation of the images generated in the workshop. In general, participants felt the output images informed and interesting.

Easy to Generate Atmosphere

“ The reason why I think the background of the generated picture looks good: it may be that when people pay attention to the painting, they do not pay special attention to the mechanism of creating a background with a sense of atmosphere. There won't be a lot of accurate detail on it, because in fact, ordinary people don't know much about the background texture perspective of painting. ”

Background, From Miss. C

“ There is an iron the railing...? But when you zoom in on a whole piece of the subject, you'll also find some places that don't quite make sense. ”

Background, From Miss. C

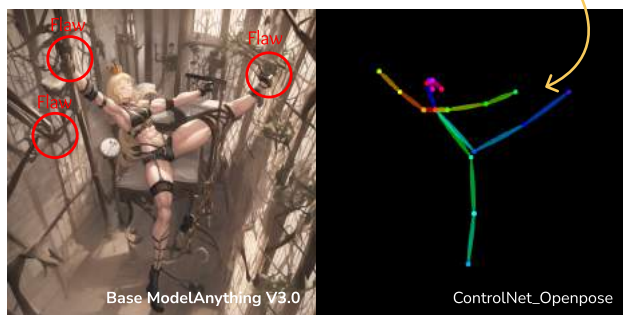


Figure 4.2-1 Image 1 Positive Prompt: masterpiece, best quality, muscularity, light, real, eyes_closed, blonde_hair, gothic,crown, armband, ankle_lace-up
Negative Prompt: lowres, bad anatomy, bad hands, text, error, missing fingers, extra digit, fewer digits, cropped, worst quality, low quality, normal quality, jpeg artifacts, signature, watermark, username, blurry, bad feet

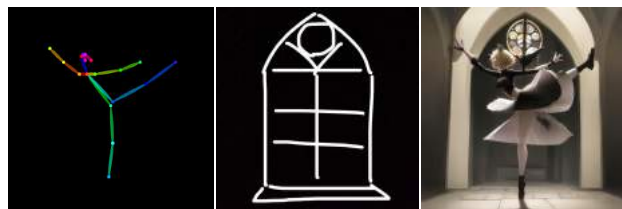


Figure 4.2-2 Image 2

“ It's pretty good. For example, the window in my image generated, I gave it such a shape, it has actually been drawn almost in the second time. ”

ControlNet_Effect, From Miss. L

Conflicts between Prompts & ControlNet

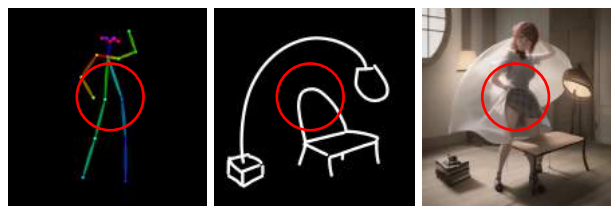


Figure 4.2-3 Image 3

“ It didn't align the tables behind the person, and it didn't know if it was a table or something else. Then the words on the left seem to be a hand book in the hands of the person, but when it comes down, he becomes a shelf. ”

ControlNet_Conflicts, From Miss. L

Training Image Dataset brings “Culture Gap”



Figure 4.2-4 Image 4

“ Really! all described in English-speaking culture. AI has a cultural generation gap? It couldn't identify the particularly famous auditorium in China as The graffiti drawing. However, people influenced by the culture will see it at a glance, but AI cannot recognize it in English. ”

Dataset Scope, From Mr. D

Pick The Right Base Model!

“ It seems this AI is better at generating 2D anime style. Is it because the hook lines in the two-dimensional anime are relatively clear? There is a clear boundary. As for generating realistic style, sometimes AI doesn't know where his boundary is, and so on. And the 2D anime itself is the product of abstract, is it just the simplified style extracted from it. ”

Style, From Miss. Y

Ideas for Application

Here are the brainstorming ideas concerning where and how can the AIGC tool help with design process from all the participants.

Design Stage

Good at Inspiration

The text-to-image method, stable diffusion, is good at generating inspiration images. This method enables rapid digitization of hand-drawn sketches and the creation of images from textual descriptions, leading to vast applications in various fields such as fashion, design, architecture, and more. In particular, SD's ability to generate high-quality inspiration images for various creative fields has been noted [1]. With its exceptional flexibility and customization capabilities, stable diffusion could become an exciting, rapidly evolving technology ushering in a new era of creative possibility for artists, designers, and creatives of all kinds.

Challenging To Control the Image Segments

Despite its potential, beginners without specific guidance may find it challenging to use prompts to control details perfectly in stable diffusion, even with the help of ControlNet. This is because generating a perfect final design effect with stable diffusion requires a significant understanding of how to manipulate different parameters and prompts to achieve the desired result. Moreover, the generated images may have some minor distortions or inconsistencies that might require many rounds of tweaks to perfect. So it is challenging at delivering perfect final design effect.

“

The AIGC model now can easily generate a painting in Van Gogh style, however, a zero to one creation can be hard for it, such as creating a Doraemon Character from 0.

So I will feel that the current AI capabilities are more to assist designers to do some uh, inspiration, rather than as a real output.

”

Application_Design Process, From Miss. L

Potential Application Fields Mentioned in Brainstorming Workshop

Personal Settings

"Personal Setting" refers to the creation and customization of a digital character or avatar according to an individual's own preferences and specifications. These specifications can range from physical features such as hair, eye color, and body shape to additional attributes such as animal characteristics or clothing choices. Users can create, customize, and raise a digital character with qualities they desire and find an artist to draw the character according to their unique specifications. The process of creating and refining the digital character can be likened to the feeling of creating a personalized anime digital twin.

And stable diffusion can greatly enrich the possibilities of Personal Settings' output.

“

Put the keywords of the personal setting features they hate into the "negative prompt". The AI tool will generate the images they want more. Maybe they won't use the AI results directly, but it will provide them with a lot of inspirations.

”

Application_Field, From Miss. L

Game Original Characters Design

Text-to-Image AI has been applied to designing original characters in games due to its ability to generate high-quality images with multiple elements that can be drawn by the input text.

“

Now our company already has an AI-generated model for game character design, and our team is optimizing the details for our game group.

”

Application_Field, From Miss. X

Potential Application Fields

Clothing Design

Stable Diffusion AI is rapidly gaining prominence in the fashion industry as an innovative tool for creating highly customized clothing designs. This AI technology is particularly beneficial for fashion designers looking to create unique and original pieces for their clients. The advanced algorithmic approach used by Stable Diffusion draws from a vast training atlas, which consists of an enormous number of clothing and fabric materials. As a result, the images generated by stable diffusion possess an impressive diversity of styles and patterns, making them an excellent choice for fashion designers looking to produce on-demand looks that stand out in the industry.

Moreover, stable diffusion's exceptional capability for customization enables fashion designers to easily manipulate and adapt the generated images to suit their specific requirements. As a result, designers have a vast array of options for creating unique garments that meet different styles and sizes.

The use of stable diffusion AI in clothing and fashion design underscores the rapid evolution of technology and its potential to revolutionize the fashion industry. With this innovative tool, designers can push the boundaries of fashion, inspire new trends, and create unforgettable new looks that reflect the unique vision of a designer, establishing new possibilities for creating clothing that is creative, unique, and tailored to each individual's needs and preferences [1].



My cousin is Z generation. I found that girls like Lolita skirts. Due to limited financial means, my cousin would use the iPad to draw some Lolita skirts to satisfy her imagination. This AI can be useful.



Application_Field, From Miss. X

Conclusion: inspiration + clothing

Develop Stable Diffusion AI as an inspiration tool can be a more achievable plan. Based on the prompt control, it is better to add more control methods to let beginners quickly generate design ideas in early design process without much knowledge of prompt engineering.

“Clothing Design” can be a promising track. Because both “Personal Settings” and “Game Original Character Design” include Clothing Design. The quality of the clothing generated significantly influence the visual effect of “Personal Settings” and “Game Original Character Design”.

4.2 Reference

[1] Grabe, I., & Zhu, J. (2023). Towards Co-Creative Generative Adversarial Networks for Fashion Designers. arXiv preprint arXiv:2304.09477.

4.3 Insights Combined With User Research

Two Basic Points

DINED Core Advantage-Body Shape

The DINED tool is a highly professional anthropometry tool with access to a comprehensive human body database. Generating anthropometry 3D mannequins is its core capability that users have found extremely valuable. In other words, While DINED's current data input method may not be an optimal interaction experience and there might be convenient AI solutions , it does not prevent users from using the 3D results generated by DINED. Thus, body shape is an important aspect to consider in future design improvements.

AI Method: Stable Diffusion

Considering the author's personal ability and the realizability of the AI algorithms, the author chose Text-to-Image track which is relatively easy to realize the prototype. Considering the characteristics of different AI models and products, the author chose stable diffusion. Because stable Diffusion stands out for its exceptional flexibility and customization options. There are numerous model versions available that enable users to not only create uncensored images, but also generate high-quality images modeled directly from Midjourney images. Such extensive customization options make Stable Diffusion unique in its class.

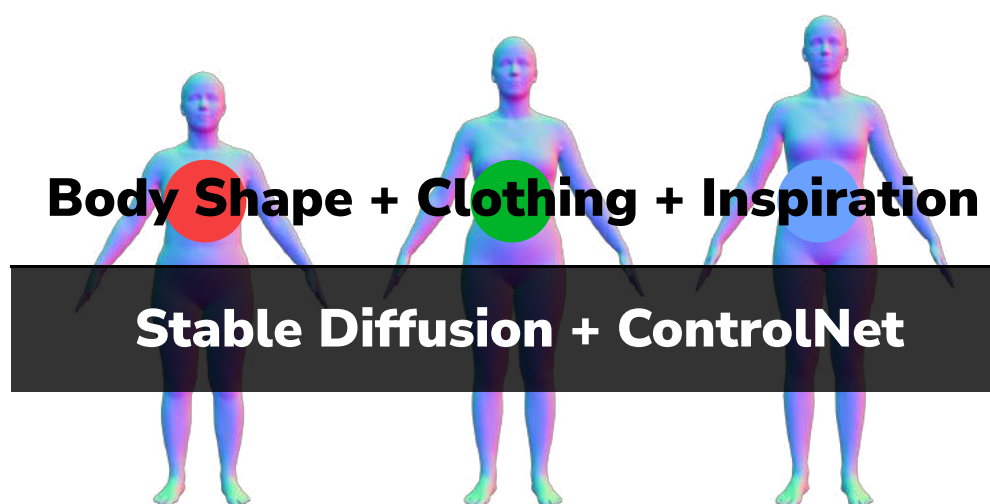


Figure 4.3-1 Formula of Concept

4.4 Final Concept

Design for Diverse Body Shapes with AI

What is Diverse Body Shape?

Diverse Body Shape refers to the ways in which different body shapes and sizes are considered inclusively and aesthetically pleasing in various contexts. This can include the cultivation of somatic aesthetic experiences, forms of aesthetic self-constitution, intentional reshaping of our aesthetic judgments of bodies, and practices of using the body aesthetically to achieve moral [1]. Body aesthetics can be a source of delight for both the subject and the object of the gaze [2]. However, aesthetic consideration of bodies also raises acute ethical questions: the body is intertwined with identity and sense of self, and aesthetic assessment of bodies can perpetuate oppression based on race, gender identity, sexual orientation, age, size, and disability [2].

Why for Diverse Body Shape ?

First and foremost, after conducting exploratory interviews, it becomes clear that a significant number of DINED users have expressed specific concerns regarding the design aspects related to diverse body shapes. This particular issue aligns with the fundamental purpose of the DINED tool as an anthropological resource.

The body shape plays a vital role in determining the right clothing options. However, existing clothing recommendation approaches often rely on a generic "one size fits all" methodology. Unfortunately, these techniques and datasets fail to consider the diverse range of body types, thus limiting the inclusivity of fashion. As a result, individuals with different body shapes do not receive suitable clothing recommendations, which undermines the potential for fashion to embrace inclusiveness.

It is crucial to acknowledge that every body is beautiful and unique, but it is only one aspect of a person's identity. [3]

4.3 Reference

[1] <https://academic.oup.com/book/26175/chapter/194275484> Irvin, Sherri (ed.), 'Introduction: Why Body Aesthetics?', in Sherri Irvin (ed.), *Body Aesthetics* (Oxford, 2016; online edn, Oxford Academic, 18 Aug. 2016), <https://doi.org/10.1093/acprof:oso/9780198716778.003.0001>, accessed 2 June 2023.

[2] Irvin, Sherri (ed.), *Body Aesthetics* (Oxford, 2016; online edn, Oxford Academic, 18 Aug. 2016), <https://doi.org/10.1093/acprof:oso/9780198716778.001.0001>, accessed 31 May 2023.

[3] <https://www.medicalnewstoday.com/articles/body-types>

4.5 Example Case

Case Topic: Design New Style Woman Wetsuit for Diverse Body Shape Aesthetic (NL)

Target Group: Anyone who wants customized wetsuit design

Case Design Task OKR:

Object: Customized Design- a wetsuit that fits the Generation Z needs

Key results 1 (Physical)

Anthropometry: suitable measurements

fitness of motion (material selection and garment cutting design, set aside)

Key results 2(Psychological)

social: good-looking

aesthetic: diverse body shape aesthetic

Design Strategy:

1.The design needs to have the aesthetics of different body shapes, with a sense of design and fashion;

2.The product fits the body size of a young Dutch woman and is able to meet the athletic needs of surfing;

Wetsuit Designer Workflow:

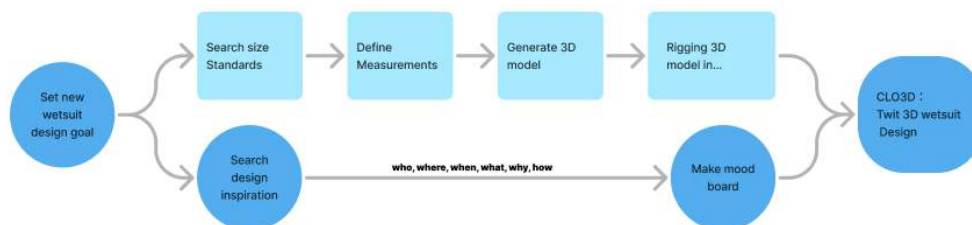


Figure 4.5-1 Wetsuit Designer Workflow Chart the light blue according to one participant in Explorative Interviews

Figure 4.5-2 Wetsuits for Different Body Shape



Figure 4.5-3 Use Dined Mannequin Tool to Generate Different Body Shapes



4.6 Design Goal

***My design goal is to **empower**
DINED users to take their first
steps in diverse body shape
related design with **AI*****



My design goal with DINED AI (DINED/Mannequin Tool with AI abilities) is to

1. **Establish it as the go-to solution for novice AI tool users in DINED users.** Because this technology became open source, personal free deployment still poses certain challenges, and the market for real body data design products is virtually nonexistent. As a result, the majority of DINED users are likely to be AI novices.
2. **Let users feel guided, easy, and elevated during the initial AI learning stage.**
3. **Attract more users in the traditional profession to try on designing with text-to-image AI ability.**

Image Source: <https://www.istockphoto.com/nl/foto/ander-ras-het-beeld-van-de-diversiteit-en-het-portret-van-de-groottevrouwen-gm1225596371-360801298>

Chapter 4

Reference (9)

Chapter 3

3.1 Reference

[1] Chou, J. (2022, December 16). Flow Chart - Deerlight Design - Medium. Medium.<https://medium.com/deerlight/design-755bfbe988b9>

3.2 Reference

[1] A. M. Hickey and A. M. Davis, "Elicitation technique selection: how do experts do it?," Proceedings. 11th IEEE International Requirements Engineering Conference, 2003., Monterey Bay, CA, USA, 2003, pp. 169-178, doi: 10.1109/ICRE.2003.1232748.

[2] Frow, P., Nenonen, S., Payne, A., & Storbacka, K. (2015). Managing co-creation design: A strategic approach to innovation. *British journal of management*, 26(3), 463-483.

Chapter 4

4.1 Reference

[1] Müller, M., & Pfahl, D. (2008). Simulation methods. *Guide to advanced empirical software engineering*, 117-152.

[2] Lee, J. H., & Ostwald, M. J. (2022). The relationship between divergent thinking and ideation in the conceptual design process. *Design Studies*, 79, 101089.

[3] Andrew. (2023, May 15). Stable Diffusion prompt: a definitive guide - Stable Diffusion Art. *Stable Diffusion Art*. <https://stable-diffusion-art.com/prompt-guide/>

4.3 Reference

[1] <https://academic.oup.com/book/26175/chapter/194275484> Irvin, Sherri (ed.), 'Introduction: Why Body Aesthetics?', in Sherri Irvin (ed.), *Body Aesthetics* (Oxford, 2016; online edn, Oxford Academic, 18 Aug. 2016), <https://doi.org/10.1093/acprof:oso/9780198716778.003.0001>, accessed 2 June 2023.

[2] Irvin, Sherri (ed.), *Body Aesthetics* (Oxford, 2016; online edn, Oxford Academic, 18 Aug. 2016), <https://doi.org/10.1093/acprof:oso/9780198716778.001.0001>, accessed 31 May 2023.

[3] <https://www.medicalnewstoday.com/articles/body-types>

05

Implementation of Functions



Chapter 5 delves into the implementation of the product's backend functionality through the utilization of both the ComfyUI GUI and the Kohya_ss GUI. This chapter also presents valuable insights gained from training a LoRA model through the examination of two specific cases.

Chapter Overview

5.1 ComfyUI

5.2 ControNet in ComfyUI

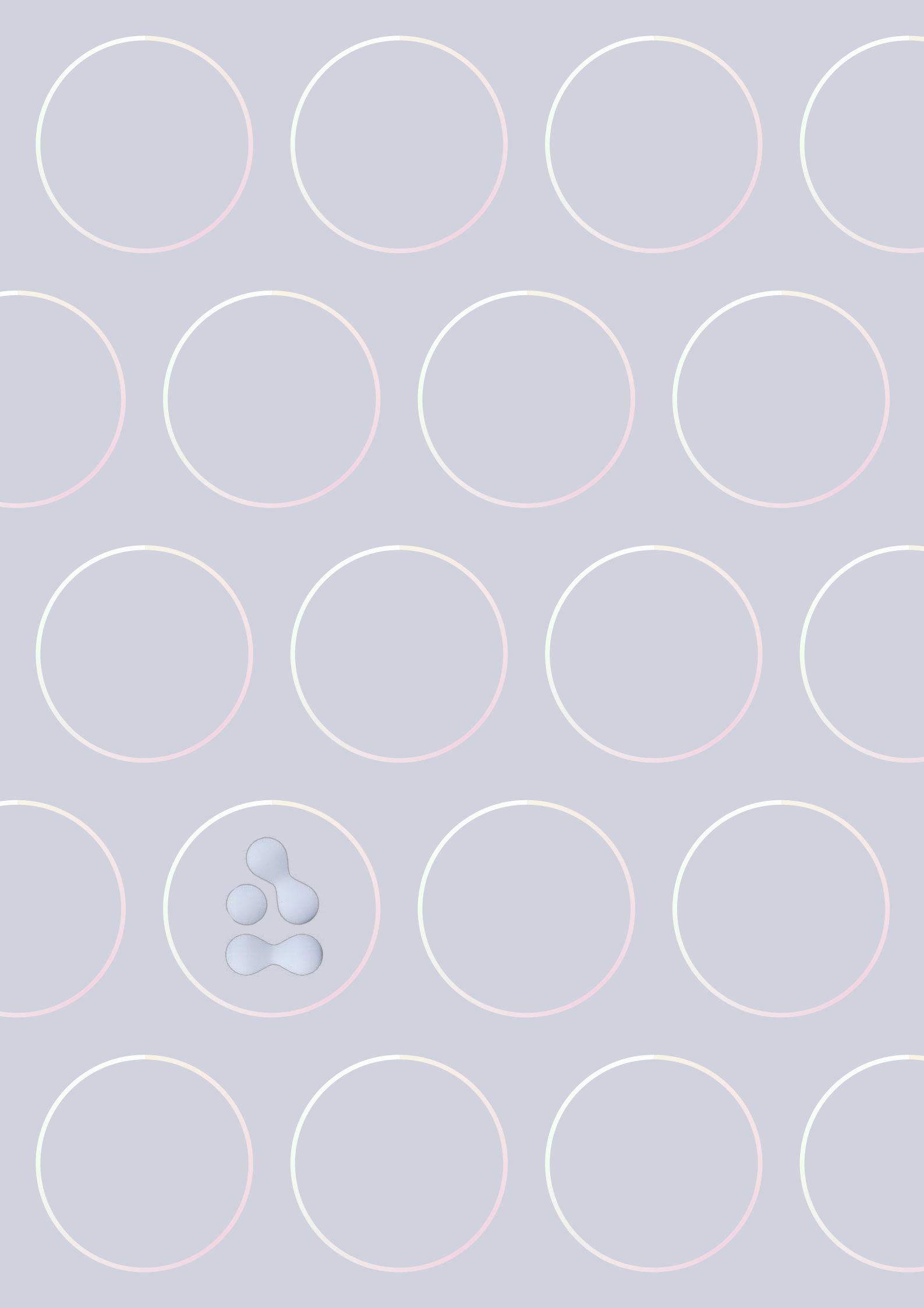
5.3 LoRA Model

5.4 Kohya_ss , A GUI for Training LoRA Models

5.5 Case1 : Efficiently Generate Inspired Visuals

5.6 Case2 : Efficiently Generate Inspired Visuals

Reference



5.1 ComfyUI

ComfyUI is a powerful and modular stable diffusion GUI with a node-based interface that allows users to design and execute advanced stable diffusion pipelines. It supports SD1.x and SD2.x and offers many optimizations, making it a popular choice for users working with stable diffusion. Some users have found ComfyUI to be more convenient and user-friendly than other stable diffusion UIs, such as SDUI [1]. This is because ComfyUI's node-based interface allows for better visualization and control over the AI image generation process. In this project, Comfy UI is the main producing AIGC tool [2].

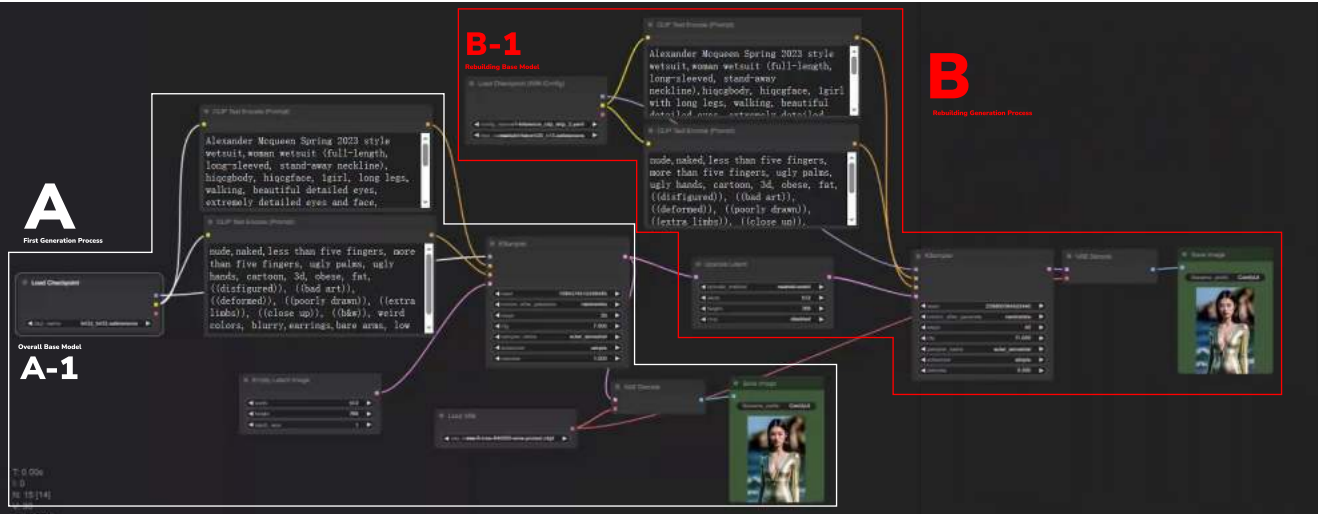


Figure 5.1-1 Two Round Generation Flow in ComfyUI (Upscale Latent) [A]_First Generation Process / [B]_Rebuilding Generation Process (Upscale Latent Process) / [A-1]_Overall Base model / [B-1]_Rebuilding Base Model

Two-Round Generation Flow :

Two round generation flow is the most commonly used flow in this project. One of the advantages of this flow is that it provides another chance to generate images of good quality via second generation. Usually, the prompts of First Generation Process and Rebuilding Generation Process should keep consistent. The output image in first generation process is input as latent image into Rebuilding Generation Process. Though it is not destined to gain a better result in rebuilding generation process. This flow can always offer some nuance changes with a base model different from the overall base model. In addition, the difference between the generated results and the training atlas can be increased by mixing different base models, thereby avoiding potential infringement disputes.

In Figure 5.1-2, it is evident that the facial features in the second image undergo a noticeable transformation, with her eyes exhibiting a remarkable level of realism due to the utilization of the Realistic Vision 2.0 base model. Despite these enhancements, the overall fashion ambiance in the image remains captivating and well-preserved.



Figure 5.1-2 Changes Between First Generation and Rebuilding Generation [Left]_Output in First Generation Process / [Right]_Output in Rebuilding Generation Process (Upscale Latent Process) /

Visual Area Conditioning / Latent composition

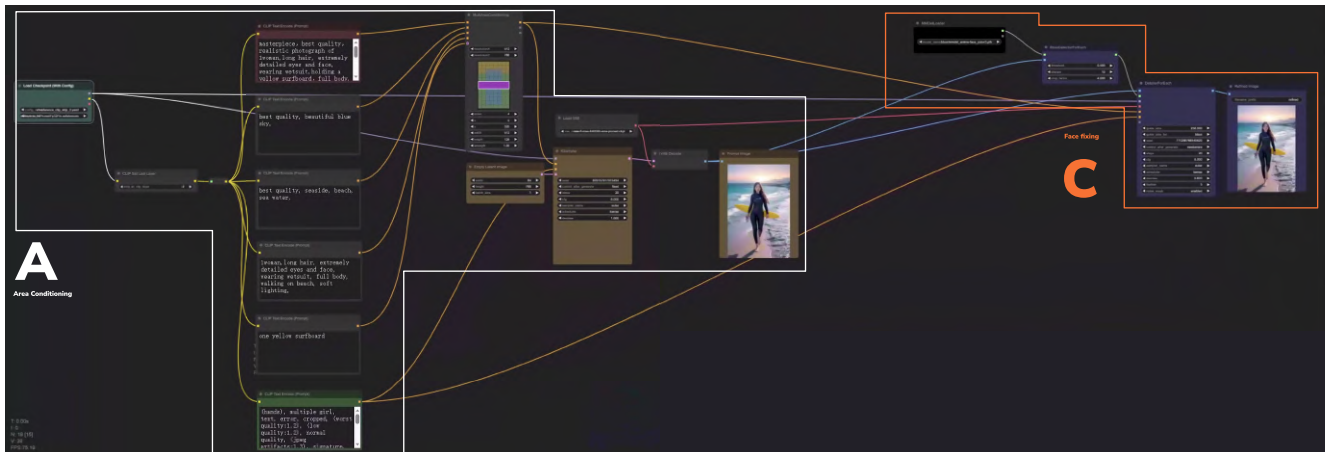


Figure 5.1-3 Layout Setting Generation Flow in ComfyUI [A]_Area Conditioning / [C]_FaceDetailer (old version)

Layout Setting Generation Flow

The A process is **Davemane42's Custom Node for ComfyUI** [3] (https://github.com/Davemane42/ComfyUI_Dave_CustomNode). This functionality empowers users to indicate specific area and assign a specific prompt to it. This function is based on a square of 64*64 pixels to generate an area unit. The current accuracy of this function's positioning is not optimal. It is crucial that the subdivision prompts are coherent and consistent with the overarching prompts.

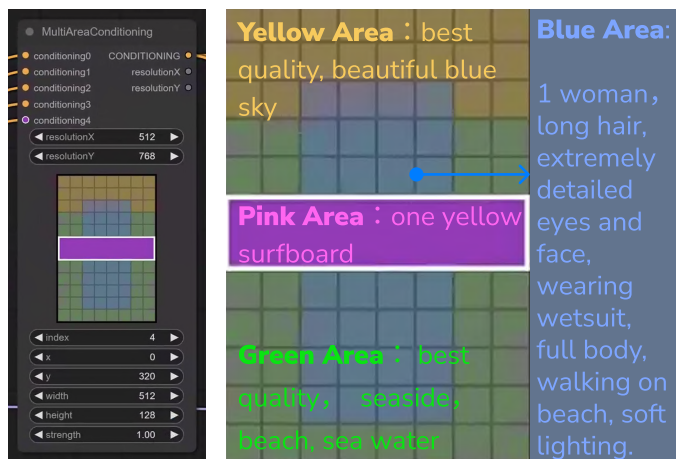


Figure 5.1-4 MultiAreaConditioning Node & Prompts in Each Color Segment [Left]_Main nodes [Right]_Area Settings



Figure 5.1-4 Face Changes [Left]_Output in First Generation Process / [Right]_Output in Face Fixing Process

Face Fixing

The B process is FaceDetailer enhancing function in ComfyUI-Impact-Pack [4] (<https://github.com/ltldrdata/ComfyUI-Impact-Pack>). To restore the details of a damaged face caused by low resolution, the FaceDetailer node utilizes high-resolution generation and synthesis techniques. This node combines a Detector node for face detection and a Detailer node for image enhancement. Together, they work to detect faces and enhance the quality of the images, effectively restoring the fine details.

5.1 Reference

- [1] Stable Diffusion Online. (n.d.). <https://stablediffusionweb.com/#demo>
- [2] Comfyanonymous. (n.d.). GitHub - comfyanonymous/ComfyUI: A powerful and modular stable diffusion GUI with a graph/nodes interface. GitHub. <https://github.com/comfyanonymous/ComfyUI>
- [3] Davemane. (n.d.). GitHub - Davemane42/ComfyUI_Dave_CustomNode. GitHub. https://github.com/Davemane42/ComfyUI_Dave_CustomNode
- [4] Ltrdrdata. (n.d.). GitHub - ltrdrdata/ComfyUI-Impact-Pack. GitHub. <https://github.com/ltrdrdata/ComfyUI-Impact-Pack>

5.2 ControlNet in ComfyUI

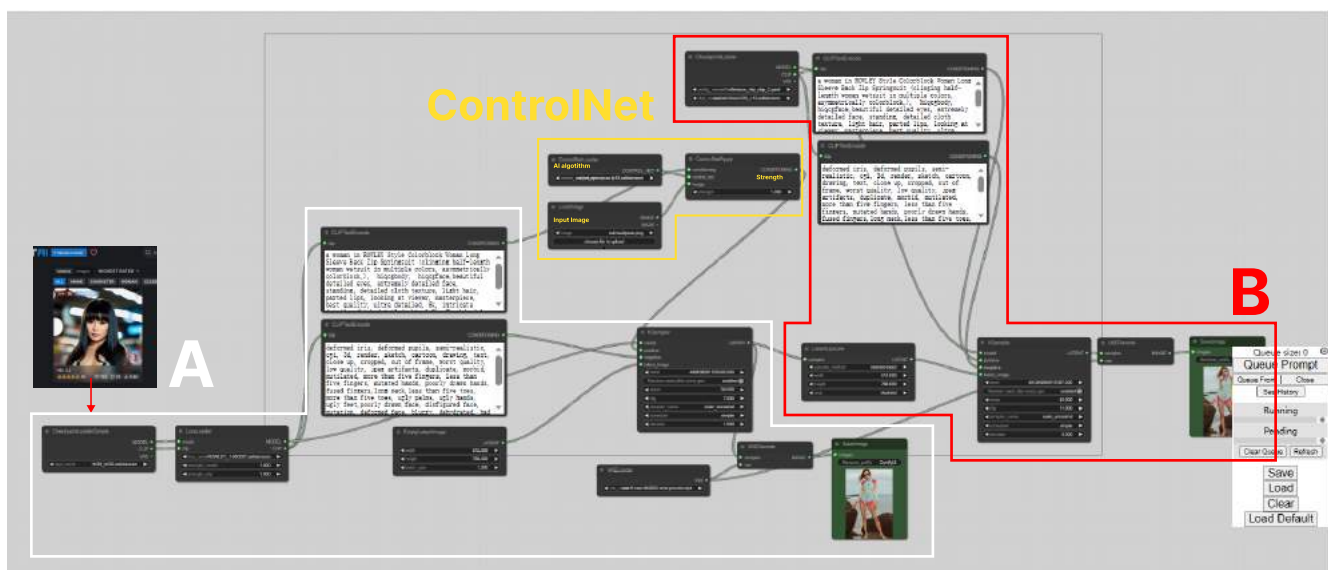


Figure 5.2-1 ControlNet in ComfyUI

There is a ControlNet modular in ComfyUI GUI, which consists of three nodes including **ControlNet_Loader**, **Load_Image** and **ControlNet_Apply**.

At **ControlNet_Loader**, the model type (open_pose, canny, depth and so on) of ControlNet can be selected.

At **Load_Image**, users should upload the image with intended control. Different types of ControlNet requires different kinds of images.

At **ControlNet_Apply**, users can employ further visual guidance to the diffusion model via the strength parameter [1]. It can be useful for fine-tuning the generated output based on the desired emphasis or attention [2].

It's important to note that the specific effects of changing these values may vary depending on the control model being used. Different control models may have different behaviors and responses to weight and guidance strength values. It's recommended to experiment and test different values to see what gives you the desired results [3].

5.2 Reference

- [1] Apply ControlNet - ComfyUI Community Manual. (n.d.). <https://blenderneko.github.io/ComfyUI-docs/Core%20Nodes/Conditioning/ApplyControlNet/>
- [2] Mikubill. (n.d.). GitHub - Mikubill/sd-webui-controlnet: WebUI extension for ControlNet. GitHub. <https://github.com/Mikubill/sd-webui-controlnet>
- [3] Ateist. (n.d.). Reddit - Dive into anything. https://www.reddit.com/r/StableDiffusion/comments/117ujqu/whats_the_difference_between_weight_and_guidance/
- [4] Llyasviel. (n.d.). GitHub - llyasviel/ControlNet: Let us control diffusion models! GitHub. <https://github.com/llyasviel/ControlNet>

Model Types Relevant To 3D Body Shape In ControlNet

Open_Pose Model:

Open_Pose Model can generate a portrait with similar pose as the input image. Open_Pose Model requires skeleton image as an input. In [lllyasviel](https://github.com/lllyasviel), an independent GUI for ControlNet, there is an embedded algorithm turning a photo/portrait into skeleton image. However, in ComfyUI, users need to upload skeleton images directly.

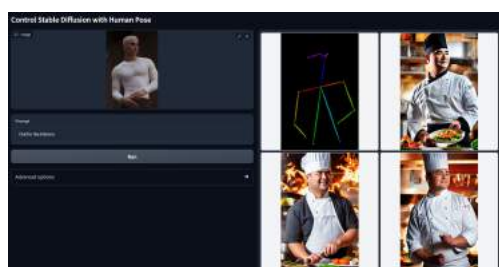


Figure 5.2-1 Open_Pose Examples [4] (<https://github.com/lllyasviel/>)

ControlNet

Depth Model:

A depth map is a 2D image in which each pixel records the distance from the viewpoint to the surface of the occlusion (which is a shadow generated object), and the corresponding vertices of these pixel are “visible” to the observer. [1]

Depth Model can generate different style portrait with the similar depth information according to depth map. Depth Model requires depth map as an input. Depth map controls the whole image (both objects and background). In ComfyUI, users need to upload depth map directly.

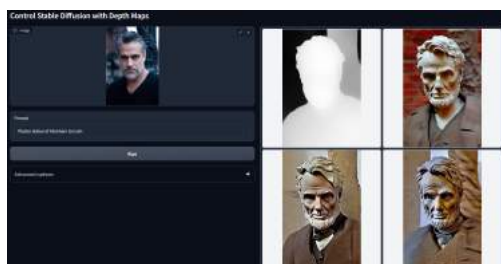


Figure 5.2-3 Depth Examples [4] (<https://github.com/lllyasviel/ControlNet>)

Canny Model:

Canny Model can generate different style portrait with the similar outlines according to the contour of the image. Canny model requires canny map as an input. The more detailed the canny map is, the more similar the output image is. In ComfyUI, users need to upload canny map directly.

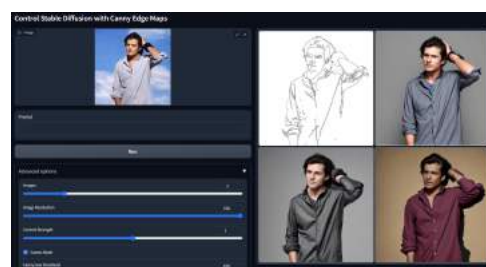


Figure 5.2-2 Canny Examples [4] (<https://github.com/lllyasviel/ControlNet>)

Normal Map:

Normal Map is a kind of Bump Map. The principle is to use the values recorded in the texture to interfere with the normal parameters in the illumination equation, so as to change the illumination result and simulate the effect of the surface fine texture. The Normal Map is stored in a normal vector and can be used directly.

Normal Model can generate different style portrait with the similar normal information according to normal map. Normal Model requires normal map as an input. The more detailed the normal map is, the more similar the output image is. In ComfyUI, users need to upload normal map directly.

5.3 LoRA Model

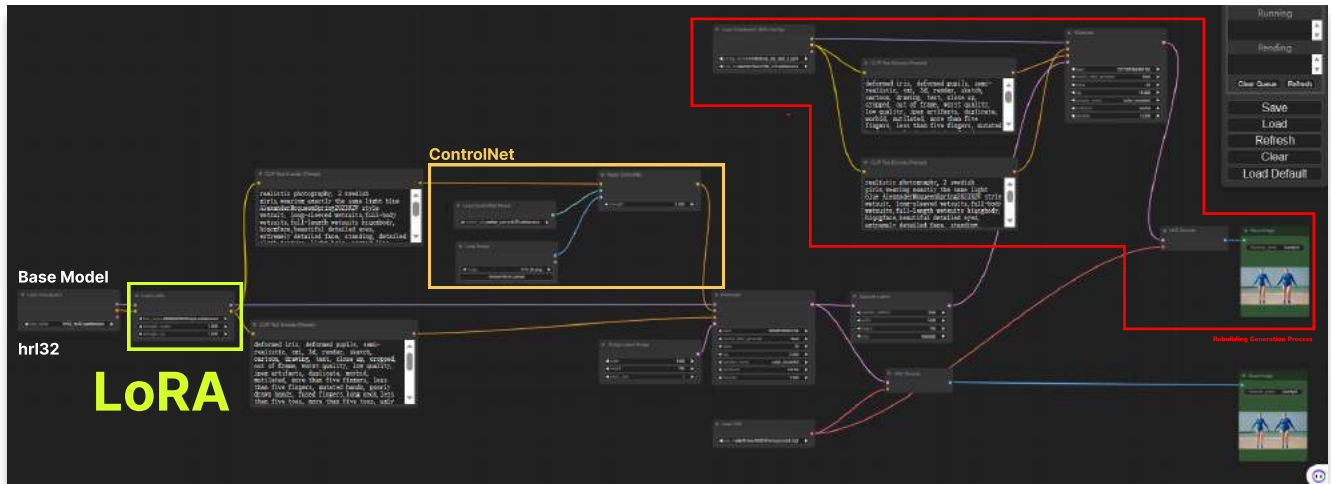


Figure 5.3-1 LoRA in ComfyUI

LoRA (Low-Rank Adaptation)[1] is a method published in 2021 that offers a way to fine-tune weights in CLIP and UNet models. These models are language models and image de-noisers employed by Stable Diffusion. LoRA enables the modification of Stable Diffusion checkpoint models' output to align with specific concepts or themes, such as art styles, characters, real-life individuals, or objects [2].

Compared to checkpoint models, LoRA models are smaller in size, typically ranging from 50MB to 1GB [3]. It's important to note that LoRA models cannot be used independently; they always require a checkpoint model to function. The usual file size of LoRA models is between 10-500MB, whereas checkpoint models span from 1 to 7GB [4]. By utilizing LoRA models in conjunction with any Stable Diffusion model, users can achieve various consistent image generation styles [5].

To employ LoRA models with the Stable Diffusion WebUI, a specific phrase needs to be included in the prompt. For instance, a prompt may appear as follows: "Prompt: (masterpiece, top quality, best quality), pixel, pixel art, bunch of red roses lora:pixel_f2:0.5" [6]. One notable advantage of LoRA models is their quick training time, which can be as little as eight minutes. Furthermore, these models have a compact size of around 5MB. This makes LoRA models faster and more memory-efficient compared to fine-tuning Stable Diffusion models using Dreambooth [7]. For those seeking free LoRA models, reputable sources include Civit.ai [8] and Huggingface.co [9]. Additionally, stable-diffusion-art.com [10] offers a range of resources for LoRA models.

In summary, LoRA models serve as small modifiers of checkpoint models, enabling efficient fine-tuning of Stable Diffusion models with reduced file sizes. They can be employed alongside any Stable Diffusion model to achieve diverse image generation styles. Platforms such as Civit.ai and Huggingface.co provide access to free LoRA models, making them readily available for utilization.

5.3 Reference (6)

[1]Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2021). Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685. [2][4] Ekman, J. (2023, June 15). LoRA In Stable Diffusion - Everything You Need to Know. Okuha. <https://okuha.com/lora/> [3][5][6]Tomeqso, & Tomeqso. (2023, May 7). How to Use LoRA Models with Stable Diffusion WebUI – Quick Tutorial - Tech Tactician. Tech Tactician - Hardware, software, guides & reviews - your own trusted tech journal. <https://techtactician.com/how-to-use-lora-models-stable-diffusion-webui/> [7][9] lora-library (lora concepts library). (2001, May 3). <https://huggingface.co/lora-library> [8] [Guide] Make your own Loras, easy and free - colabs | Stable Diffusion Other | Civitai. (n.d.). <https://civitai.com/models/22530/guide-make-your-own-loras-easy-and-free> [10] Andrew. (2023a, March 1). What are LoRA models and how to use them in AUTOMATIC1111 - Stable Diffusion Art. Stable Diffusion Art. <https://stable-diffusion-art.com/lora/>

5.4 Kohya_ss , A GUI for Training LoRA Models

Kohya_ss LoRA:

The Kohya_ss GUI allows users to set the training parameters and generate and run the required CLI commands to train the model [1].

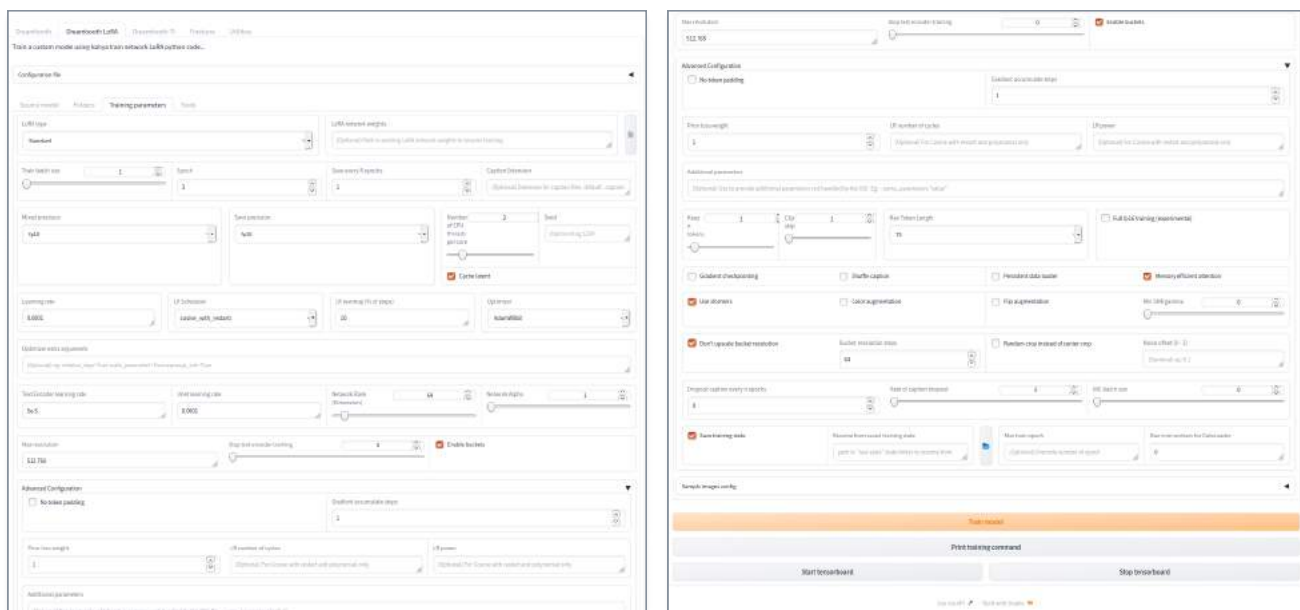


Figure 5.4-1 Screenshots of Kohya_ss GUI

Training Steps & Training Details in Kohya_ss:



Step 1

Prepare Image Datasets

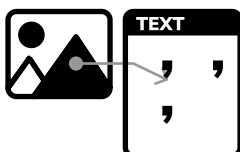
0. Select Pictures

1.Remove Background

2.Process Images

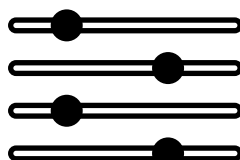
3.AI Enhance Resolution

1. To ensure optimal performance, it is crucial to have a sufficiently large image dataset. For instance, if the dataset contains fewer than 10 samples, the LoRA model becomes highly susceptible to overfitting. In such situations, one can employ various techniques like using prompts or employing inpainting methods to address this issue effectively. These approaches can help compensate for the limited data availability and enhance the model's generalization capabilities.
2. The image resolution in the dataset is of utmost importance. And it is highly recommended to maintain consistent image sizes throughout. Generally, selecting two sizes, name 512x512 and 512x768 pixels, can yield a better guiding effect. This choice is particularly beneficial because the stable diffusion basic model (e.g. SD V1.5 model) employs an image size of 512x512 [2]. By adhering to these resolutions, one can enhance the model's performance and alignment with the dataset.



Step 2

Caption & Tap the Images



Step 3

Set Parameters

3. It is preferable for the images to include elements that are similar to the desired output effect. For instance, if the goal is to generate profile photos of people, the LoRA image dataset should consist of photos featuring individuals with similar profile faces.

1. It is essential to order captions from most to least prominent concept.

2. Advice for different purpose captions:

- When training an **object LoRA model**. During the captioning process, each picture in the dataset can be uniformly marked with object keyword tags [3].
- When training a **style LoRA model**, it is important to label each picture comprehensively, including details such as decorations, clothing, background, and etc. . And one do not need to caption a style keyword which is opposite to object training.
- When extracting an **object style LoRA model**, which includes partially visible elements, from a picture dataset, such as extracting the style of a series clothing people wear, it is crucial to combine both methods mentioned above. One should caption the partial style keywords, meaning that the desired combination of elements can be succinctly summarized in a single word. Additionally, it is worth mentioning that elements that are not required in the desired output should be marked as detailed as possible.

1. Although this GUI offers numerous adjustable parameters, it is important to focus on only a few key parameters. To facilitate this, a set of parameter setting templates is recommended. Once the initial two steps have been successfully completed, fine-tuning different parameters has minimal impact. Developers should provide further explanation regarding these parameters [4].

2. The optimal parameter settings vary depending on the specific dataset. Some commonly used operational options include :

- 1). Repeating 30-100 steps
- 2). Setting epochs to 10-20.

It is possible to achieve better results when the loss value below 0.020 [5].

5.4 Reference

[1]Bmaltais. (n.d.). GitHub - bmaltais/kohya_ss. GitHub. https://github.com/bmaltais/kohya_ss

[2]runwayml/stable-diffusion-v1-5 · Hugging Face. (n.d.). <https://huggingface.co/runwayml/stable-diffusion-v1-5>

[3]Terrarium. (n.d.). Reddit - Dive into anything. https://www.reddit.com/r/StableDiffusion/comments/114dxgl/advanced_advice_for_model_training_finetuning_and/

[4]Gözükara, F. (n.d.). How to Generate Studio Quality Realistic Photos via Kohya Web GUI DreamBooth LoRA Training Full Tutorial - 35 Video Chapters - Manually Fixed Subtitle. [www.linkedin.com. https://www.linkedin.com/pulse/how-generate-studio-quality-realistic-photos-via-kohya-g%C3%B6z%C3%BCkara](https://www.linkedin.com/pulse/how-generate-studio-quality-realistic-photos-via-kohya-g%C3%B6z%C3%BCkara)

[5]Anant. (2023, April 7). Fine-tuning with LoRA: create your own avatars & styles! kix. <https://www.kix.in/2023/04/07/sd-lora-finetuning/>

5.5 Case1 : Efficiently Generate Inspired Visuals

Partly Style LoRA Model

Generate New Wetsuit from Wetsuit: This case focuses on generating design inspirations for color block style spring suits. Following the workflow outlined in this chapter, the author provides a concise overview of the dataset images, the parameter settings used for training, and presents typical outputs of the LoRA model.



Figure 5.5-1 Case 1_Image Dataset Example [1]

The dataset contains a total of 26 images featuring spring suits. From the website cynthiarowley.com[1], the author handpicked three spring suits that exhibit a similar color block style. These showcase images were chosen for their high quality and consistent size. To eliminate the influence of shadows and backgrounds, the author meticulously removed the background from each picture. Each spring suit includes both front view and back view images. The author aptly summarized the desired combination of elements as "ROWLEY Style Colorblock Woman Long Sleeve Back Zip Spring Suit" during the captioning process. This particular caption is also utilized in the ComfyUI image generation process. (Please See Appendix C5.5)

5.5 Reference

[1] Colorblock Wetsuit. (n.d.). Cynthia Rowley. <https://cynthiarowley.com/en-nl/products/color-block-wetsuit?variant=40699377156276>

[2]Sp3zisaf4g. (n.d.). Reddit - Dive into anything. https://www.reddit.com/r/StableDiffusion/comments/10ir5ax/big_comparison_of_lora_training_settings_8gb_vram/

[3]Automatic. (n.d.). GitHub - AUTOMATIC1111/stable-diffusion-webui: Stable Diffusion web UI. GitHub. <https://github.com/AUTOMATIC1111/stable-diffusion-webui>

Training Parameters	
number of dataset images	26
Resolution	512*768
batch size	1
repeat	40
Dimension (network rank)	64
Network Alpha	32
Epoches	10
lr	0.0001
LR warmup%	10%
UNET	0.0001
text	5e-5
LR Scheduler	cosine_with_restarts
steps	10400
time	03:24:07

Table 5.5-1 Parameter Settings [2]

The parameter settings mentioned above are typical for LoRA beginners, ensuring a certain degree of quality in the LoRA model while maintaining an acceptable training time on medium and low-end computer equipment (The author used a 2060ti GPU in this project). The LoRA training consists of 10 epochs, with each image in the dataset being repeatedly learned 40 times per epoch. It is recommended to set the "Dimension (network rank)" twice larger than the "Network Alpha". The larger the "Dimension value" is, the better the quality of the output image of LoRA will be. The "lr (learning rate)" is set to 0.0001. To train the model quickly, one can select "cosine_with_restarts" in the "LR (Learning Rate) scheduler" [2].

Chapter 5

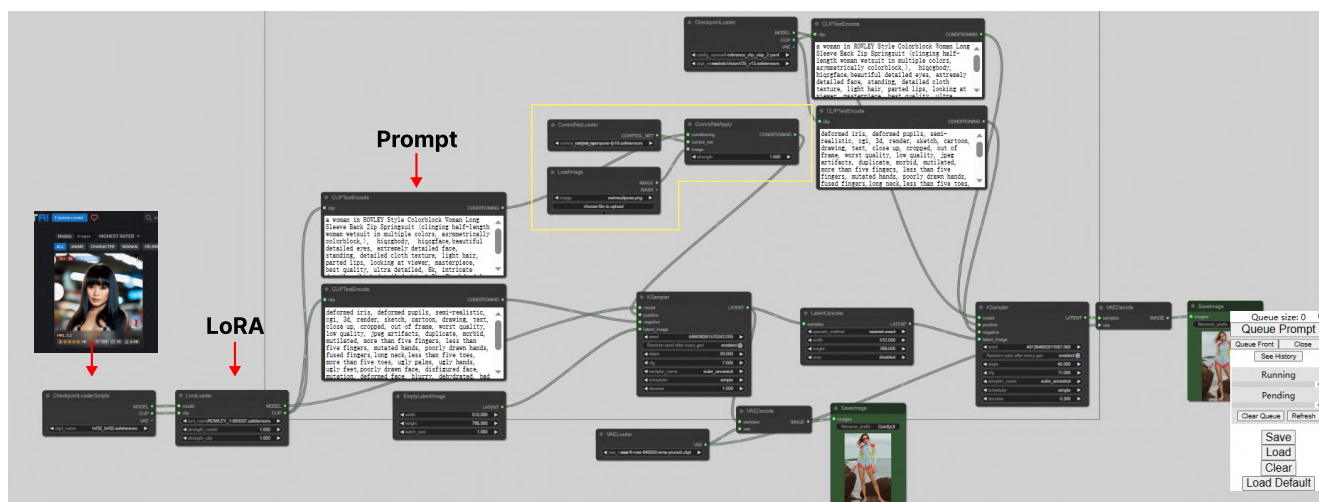


Figure 5.5-2 Case 1_ComfyUI Nodes 1.Utilize HRL 3.2 as the base model; 2. Use ControlNet_Openpose to generate images of front view and back view;
3. Two round generation process following the workflow outlined in this chapter;

Output in Each Epoch

Figure 5.5-3 shows the loss value and image output over 10 epochs when the "strength_model" is set to 1.0. The "strength_model" of the LoRA model is set to 1.0 because the XY plot test at epoch 10 with a strength of 1.0 showed a positive effect on generating image. From the perspective of generating design inspiration, the LoRA model has demonstrated a relatively good performance, particularly in terms of color contrast style and color block segmentation, which exhibit a pleasing diversity. However, there are some drawbacks, such as the lack of clarity in the generation of wetsuit details and strange effects in the generation of the back zipper strap.

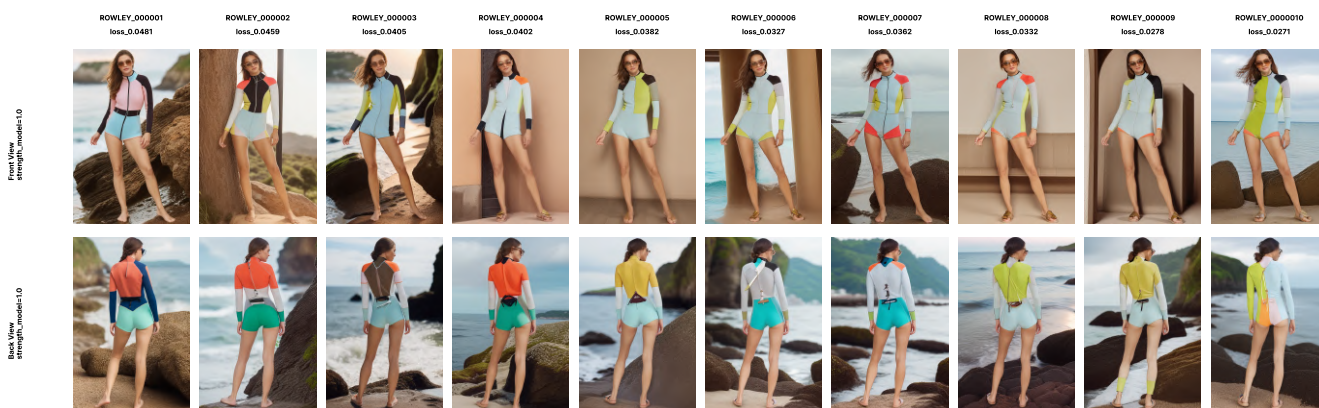


Figure 5.5-3 Case 1_Loss Value & Image Output in Each Epoch (strength_model=1.0)

Table 5.5-2, generated by the Automatica1111 GUI [3], illustrates the output image effects at different strength_model values in each epoch of the LoRA model. Two dimensions are used to assess whether the generated results are satisfactory. The first dimension examines whether the model's clothing style resembles a wetsuit (indicated by the cyan line in the table), while the second dimension evaluates whether the model's clothing style resembles a color block (indicated by the white line in the table). The portion enclosed by yellow lines represents the generated results that roughly meet both dimensions, whereas the section enclosed by purple lines corresponds to the part with the closest resemblance to the input style. Consequently, the model results from the eighth, ninth, and tenth epochs perform better.

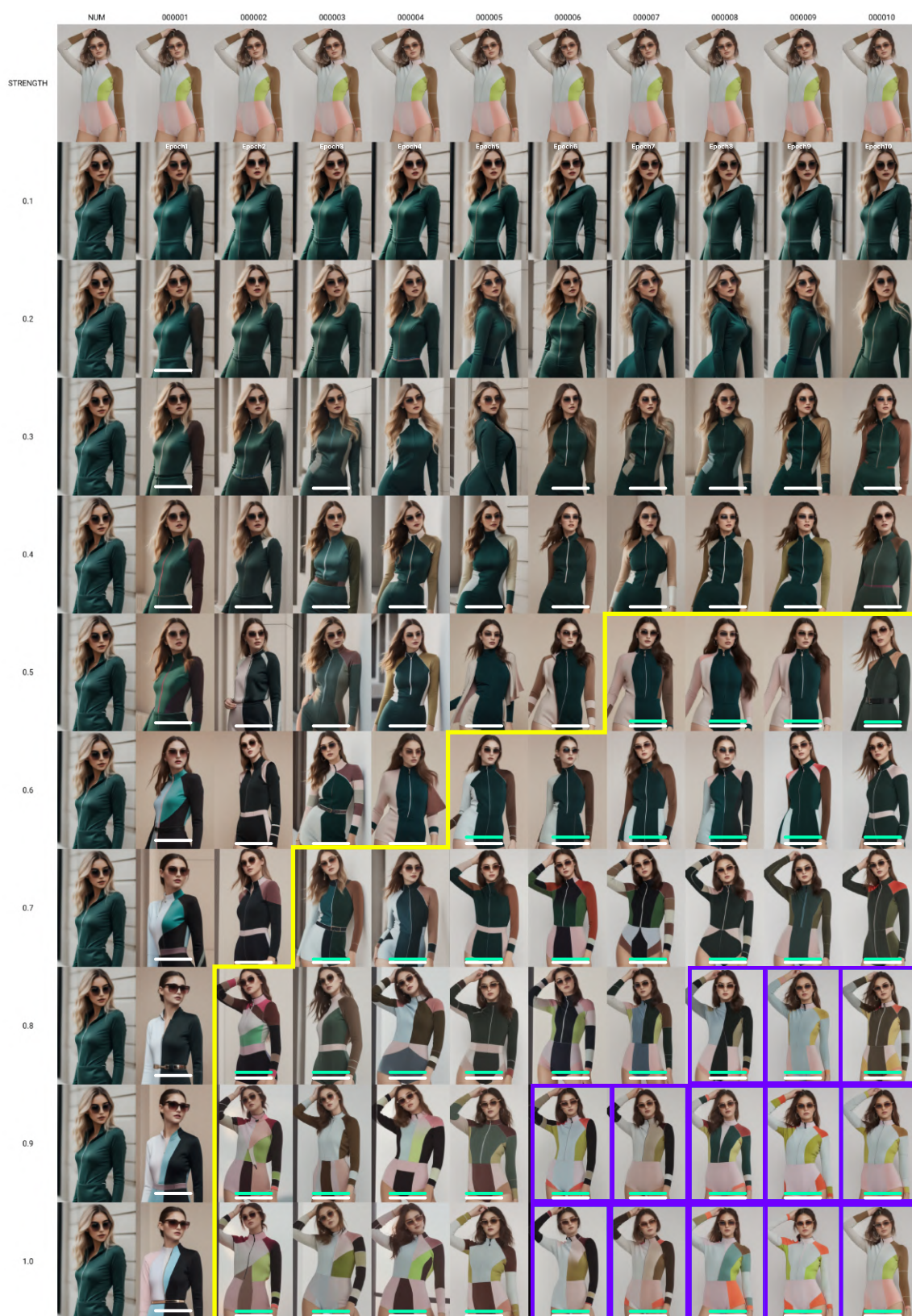


Table 5.5-2 XY Plot of Case 1 Automatica1111 If the goal is to generate color block style spring suit inspirations, it is recommended to select a LoRA model with an epoch number greater than 5 and a strength_model value greater than 0.8.

— Springsuit (wetsuit)
— Colorblock

Ideal Effect

**Roughly
Similar**

5.6 Case2 : Efficiently Generate Inspired Visuals

Partly Style LoRA Model + Style Transfer

Generate New Wetsuit from Clothing: This case revolves around the generation of design inspirations for a wetsuit with a specific style featured in the 2023 ready-to-wear Alexander McQueen spring show. The author offers a succinct summary of the dataset images, the parameter configurations employed during training, and a comparison diagram showcasing input and output results.

Since this case is to generate new wetsuit idea form clothing style, this project has a high tolerance for overfitting for that there are too few images (<10) in the dataset. However, to compensate the overfitting problem, it becomes crucial to prioritize and focus on the captions(see Figure 5.6-1).



Figure 5.6-1 Case 2_Image Dataset [2] & Typical Parameter Settings The author chose images from the Vogue Slideshow due to their high quality and consistent representation of Alexander McQueen's style.

Nodes in ComfyUI are the same in Figure 5.3-1 LoRA in ComfyUI.

1. Utilize HRL 3.2 as the base model;
2. Employ ControlNet_Canny to generate images featuring multiple women with distinct body shapes wearing the same wetsuit;
3. Implement a two-round generation process following the workflow described in this chapter;



Figure 5.4-1 Design Elements in LoRA Dataset of Alexandra Macqueen Style Wetsuit & LoRA Output

Comparison Diagram Showcasing Input and Output Results

Upon close examination of Figure 5.4-1, a striking observation emerges regarding the impressive style transferring prowess exhibited by the workflow integrated with the LoRA model in this project. This noteworthy capability, however, should not overshadow the inherent challenges posed by the intricate wetsuit processing involved. Within the realm of the LoRA Dataset, the design elements manifest themselves in the output image with remarkable finesse and precision. Notably, the harmonious integration of the chest line and waist line into the diverse range of wetsuit designs stands as a testament to the exceptional artistry at play. Such seamless amalgamation across varying sizes is truly commendable and underscores the expertise employed in this endeavor.

5.6 Reference

[1] LoRA Training Guide. (n.d.). https://reentry.org/lora_train#pruning-captions

[2] Nast, C. (n.d.). Alexander McQueen Spring 2023 Ready-to-Wear Fashion Show. Vogue. <https://www.vogue.com/fashion-shows/spring-2023-ready-to-wear/alexander-mcqueen/slideshow/collection#1>

Reference (24)

Chapter 5

5.1 Reference

- [1] Stable Diffusion Online. (n.d.). <https://stablediffusionweb.com/#demo>
- [2] Comfyanonymous. (n.d.). GitHub - comfyanonymous/ComfyUI: A powerful and modular stable diffusion GUI with a graph/nodes interface. GitHub. <https://github.com/comfyanonymous/ComfyUI>
- [3] Davemane. (n.d.). GitHub - Davemane42/ComfyUI_Dave_CustomNode. GitHub. https://github.com/Davemane42/ComfyUI_Dave_CustomNode
- [4] Ltdrdata. (n.d.). GitHub - ltdrdata/ComfyUI-Impact-Pack. GitHub. <https://github.com/ltdrdata/ComfyUI-Impact-Pack>

5.2 Reference

- [1] Apply ControlNet - ComfyUI Community Manual. (n.d.). <https://blenderneko.github.io/ComfyUI-docs/Core%20Nodes/Conditioning/ApplyControlNet/>
- [2] Mikubill. (n.d.). GitHub - Mikubill/sd-webui-controlnet: WebUI extension for ControlNet. GitHub. <https://github.com/Mikubill/sd-webui-controlnet>
- [3] Ateist. (n.d.). Reddit - Dive into anything. https://www.reddit.com/r/StableDiffusion/comments/117ujqu/whats_the_difference_between_weight_and_guidance/
- [4] Lllyasviel. (n.d.). GitHub - lllyasviel/ControlNet: Let us control diffusion models! GitHub. <https://github.com/lllyasviel/ControlNet>

5.3 Reference (6)

- [1]Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., ... & Chen, W. (2021). Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685.
- [2][4] Ekman, J. (2023, June 15). LoRA In Stable Diffusion - Everything You Need to Know. Okuha. <https://okuha.com/lora/>
- [3][5][6]Tomeqso, & Tomeqso. (2023, May 7). How to Use LoRA Models with Stable Diffusion WebUI – Quick Tutorial - Tech Tactician. Tech Tactician - Hardware, software, guides & reviews - your own trusted tech journal. <https://techtactician.com/how-to-use-lora-models-stable-diffusion-webui/>
- [7][9] lora-library (lora concepts library). (2001, May 3). <https://huggingface.co/lora-library>
- [8] [Guide] Make your own Loras, easy and free - colabs | Stable Diffusion Other | Civitai. (n.d.). <https://civitai.com/models/22530/guide-make-your-own-loras-easy-and-free>
- [10] Andrew. (2023a, March 1). What are LoRA models and how to use them in AUTOMATIC1111 - Stable Diffusion Art. Stable Diffusion Art. <https://stable-diffusion-art.com/lora/>

5.4 Reference

- [1]Bmaltais. (n.d.). GitHub - bmaltais/kohya_ss. GitHub. https://github.com/bmaltais/kohya_ss
- [2]runwayml/stable-diffusion-v1-5 · Hugging Face. (n.d.). <https://huggingface.co/runwayml/stable-diffusion-v1-5>
- [3]Terrariyum. (n.d.). Reddit - Dive into anything. https://www.reddit.com/r/StableDiffusion/comments/114dxgl/advanced_advice_for_model_training_finetuning_and/
- [4]Gözükara, F. (n.d.). How to Generate Studio Quality Realistic Photos via Kohya Web GUI DreamBooth LoRA Training Full Tutorial - 35 Video Chapters - Manually Fixed Subtitle. www.linkedin.com/pulse/how-generate-studio-quality-realistic-photos-via-kohya-g%C3%B6z%C3%BCkara
- [5]Anant. (2023, April 7). Fine-tuning with LoRA: create your own avatars & styles! kix. <https://www.kix.in/2023/04/07/sd-lora-finetuning/>

5.5 Reference

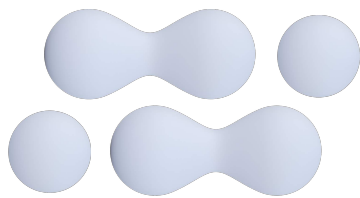
- [1] Colorblock Wetsuit. (n.d.). Cynthia Rowley. <https://cynthiarowley.com/en-nl/products/color-block-wetsuit?variant=40699377156276>
- [2]Sp3zisaf4g. (n.d.). Reddit - Dive into anything. https://www.reddit.com/r/StableDiffusion/comments/10ir5ax/big_comparison_of_lora_training_settings_8gb_vram/
- [3]Automatic. (n.d.). GitHub - AUTOMATIC1111/stable-diffusion-webui: Stable Diffusion web UI. GitHub. <https://github.com/AUTOMATIC1111/stable-diffusion-webui>

5.6 Reference

- [1] LoRA Training Guide. (n.d.). https://reentry.org/lora_train#pruning-captions
- [2] Nast, C. (n.d.). Alexander McQueen Spring 2023 Ready-to-Wear Fashion Show. Vogue. <https://www.vogue.com/fashion-shows/spring-2023-ready-to-wear/alexander-mcqueen/slideshow/collection#1>

06

DINED AI Tool Interaction Design



Chapter 6 presents the comprehensive UX and UI design developed for this project, which encompasses the entire process of LoRA model training and Text-to-Image generation. The goal is to provide one user-friendly and visually concise interface that enhances the overall user experience.

Chapter Overview

UX & UI Design

6.1 STEP_1_Foundation Type Selection

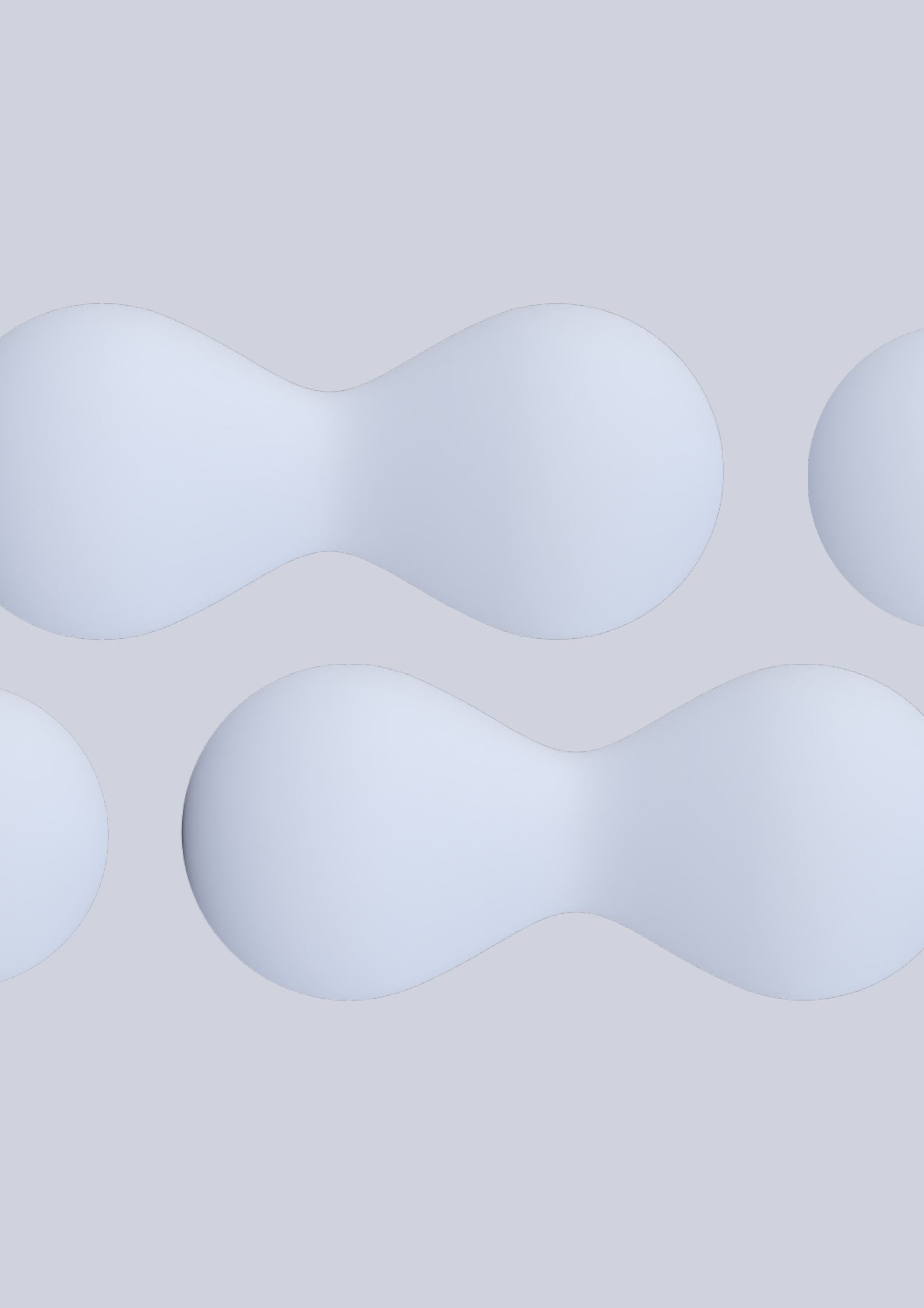
6.2 STEP_2_Object Effect Settings

6.3 STEP_3_Human Factors Settings

6.4 STEP_4_ Image Layout (Optional)

6.5 STEP_5_Final Effect

Take Away



UX & UI Design

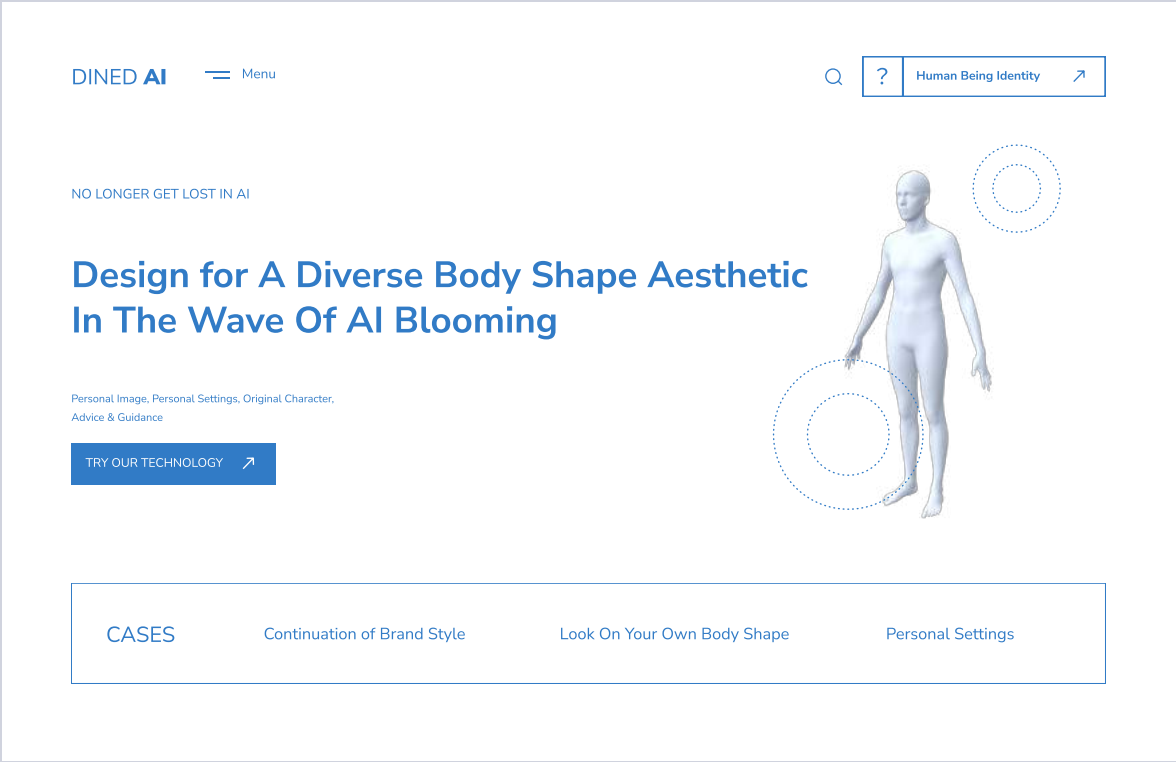


Figure 6.0-2 Home Page of DINED AI (Please See Appendix C 6.0)

Design Keynotes

In terms of UX, the author focuses on designing the product to meet the requirements of online software tools. Given that tool products aim to convey a professional atmosphere and facilitate efficient information transmission, coupled with the inherent complexity of the SD process, the author avoids excessive visual design. Instead, they strive to keep page information concise and provide users with timely and efficient guidance.

Regarding the UI aspect, the new products maintain the UI features of DINED, which include the following:

- 1. Theme Color: The product continues to utilize the DINED Blue color as its primary theme color. This consistent color scheme ensures visual continuity and allows users to readily associate the new product with DINED.
- 2. Right Angle Elements: The components within the UI incorporate right angle elements. By employing these sharp angles, the design maintains a cohesive look with DINED's established visual style.

Furthermore, the design draws inspiration from professional components, specifically referencing TDesign [1]. This approach ensures that the design is constructed efficiently and professionally, leveraging the elements and principles of professional components.

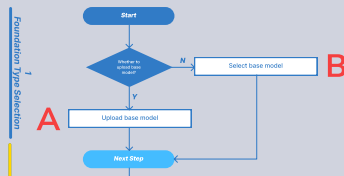
Overall, the author prioritizes the UX by focusing on efficient information delivery and providing users with effective guidance. In terms of UI, they maintain consistency with DINED's visual identity, incorporating the established theme color and right angle elements. The design also incorporates professional components, specifically referring to TDesign, to ensure a high-quality and professional construction of the product.

6.0 Reference

[1] TDesign. (n.d.-b). TDesign - 开源的企业级设计体系. <https://tdesign.tencent.com/index-en>

6.1 STEP_1_Foundation Type Selection

User Flow



Step 1 :

Foundation Type Selection

To begin, it is important to choose the appropriate foundation type of the base model, as it determines the style of the generated image. Each base model excels in generating distinct styles, such as **Realistic Vision** [1], which is renowned for its ability to produce lifelike photographs, while **Beautiful Realistic Asians** [2] specializes in generating images of beautiful young Asian girls.

Design Point 1

Integrate with Civit AI

DINED AI seamlessly integrates with the Civit AI website, allowing users to conveniently select their preferred open-source base models directly within the DINED AI tool. Furthermore, DINED AI offers the functionality to directly download and save the base models online, eliminating the need for users to navigate between different websites to acquire and upload large-size models. This streamlined approach not only saves users valuable disk space but also reduces installation time, enhancing the overall user experience of the DINED AI tool.

6.1 Reference

- [1] Realistic Vision V3.0 - v3.0 (VAE) | Stable Diffusion Checkpoint | Civitai. (n.d.). <https://civitai.com/models/4201/realistic-vision-v30>
- [2] Beautiful Realistic Asians - Brav5 | Stable Diffusion Checkpoint | Civitai. (n.d.). <https://civitai.com/models/25494/beautiful-realistic-asians>

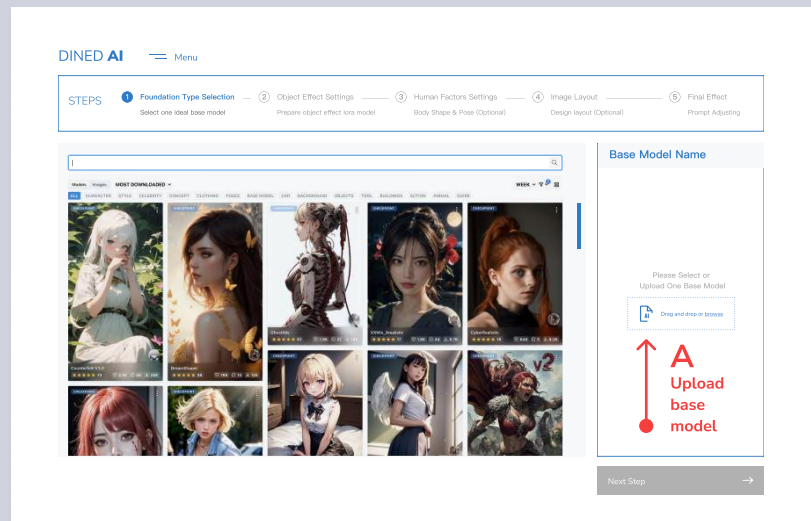


Figure 6.1-1 Upload Base Model

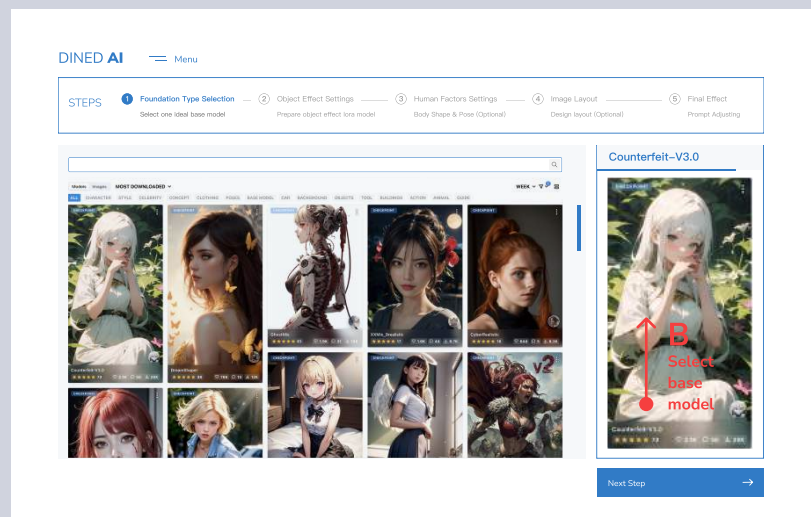


Figure 6.1-2 Select Base Model

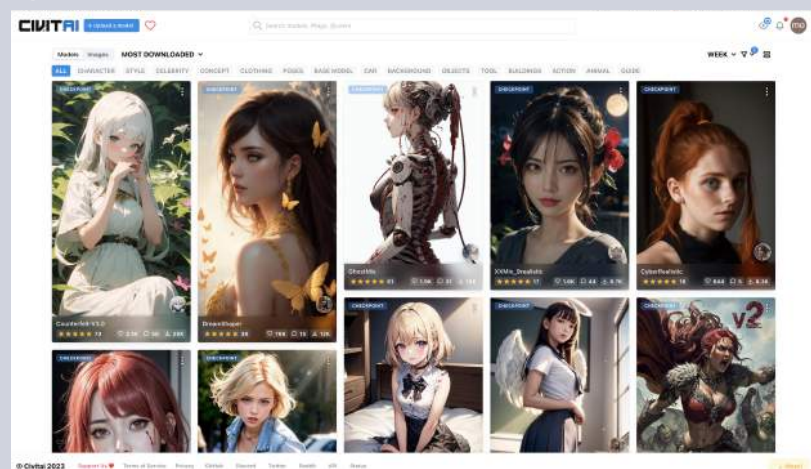
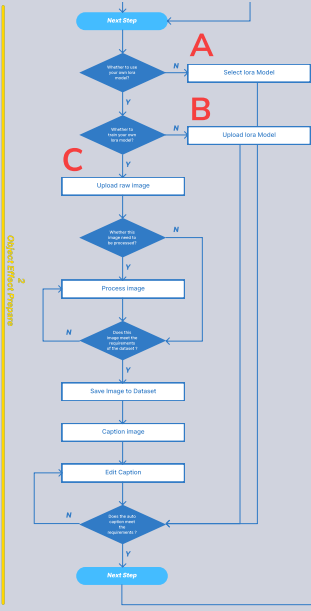


Figure 6.1-3 Civit AI Web Civit AI is a platform that focuses on facilitating the sharing of AI models and code for generating art using machine learning algorithms [3] futurehub.in. The platform encourages collaboration and sharing among artists, developers, and researchers interested in exploring the intersection of art and AI [4] futurehub.in.

6.2 STEP_2_Object Effect Settings

User Flow



Step 2 : Object Effect Settings

The next crucial step is configuring the Object Effect Settings. This particular stage empowers users to harness the capabilities of a LoRA model, which acts as a guiding force for the stable diffusion model, resulting in the generation of images with specific directional attributes. Users have the flexibility to choose from a diverse range of pre-existing LoRA models available online or train their own customized LoRA model using the comprehensive resources offered by DINED AI.

Design Point 2 Involve LoRA Model

The utilization of LoRA sets DINED AI apart from other SD tools available in the market, making it a unique and standout feature. Therefore, it is crucial not to overlook this step in the process. However, users can rest assured that they have the option to explore a wide range of LoRA models online, ensuring they can always find a model that aligns with their desired effects.

Figure 6.2-1 Object Effect Settings WebPage 1

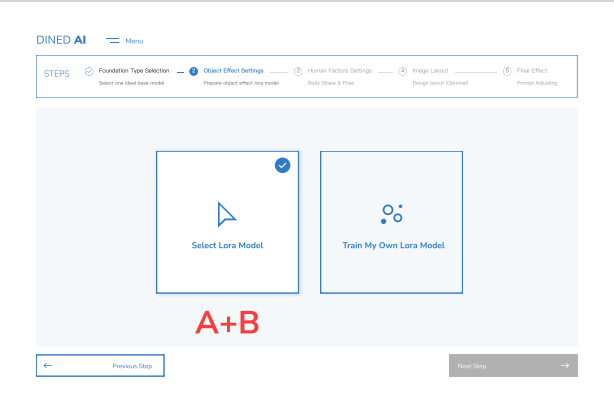
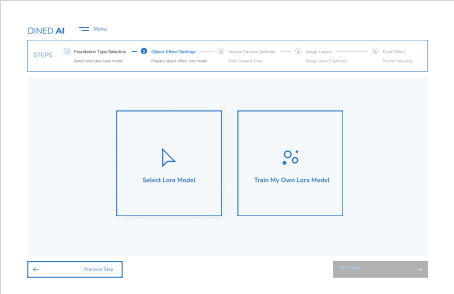


Figure 6.2-2 Select Lora Model

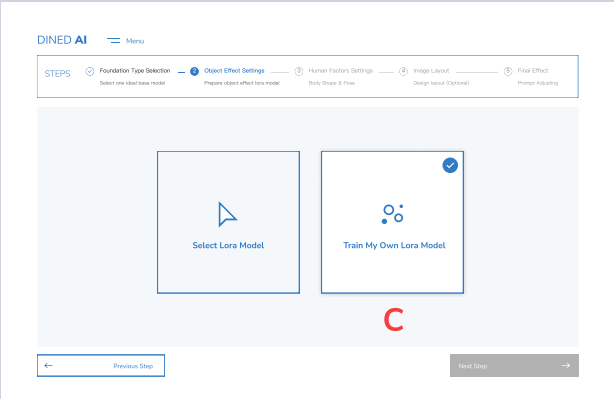


Figure 6.2-4 Train My Own Lora Model

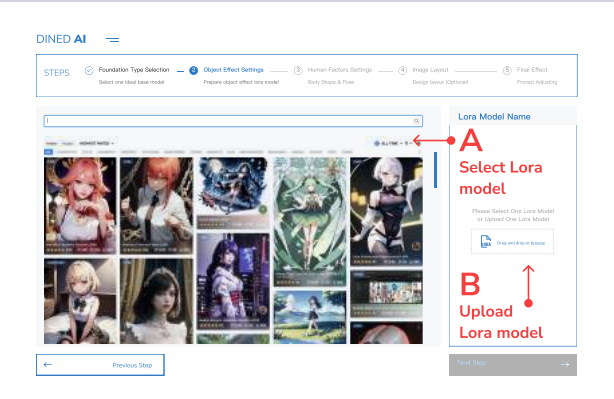


Figure 6.2-3 A-Select Lora Model B_Upload Lora Model

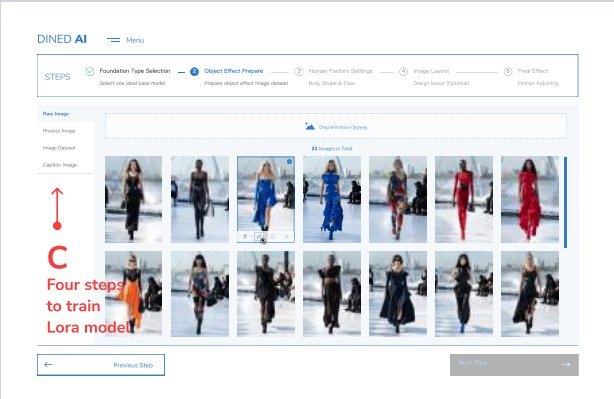


Figure 6.2-5 Four Steps to Train Lora Model

C : Object Effect Settings

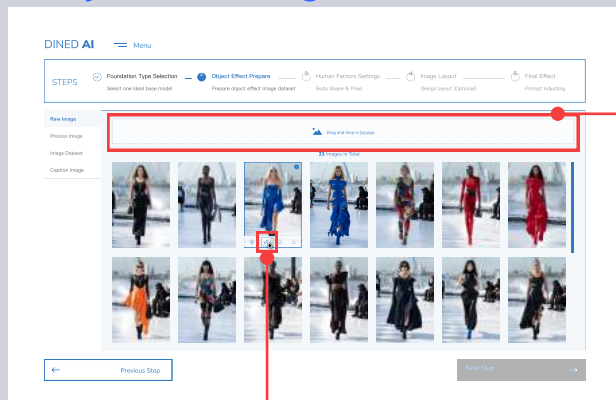


Figure 6.2.1-1 Raw Image

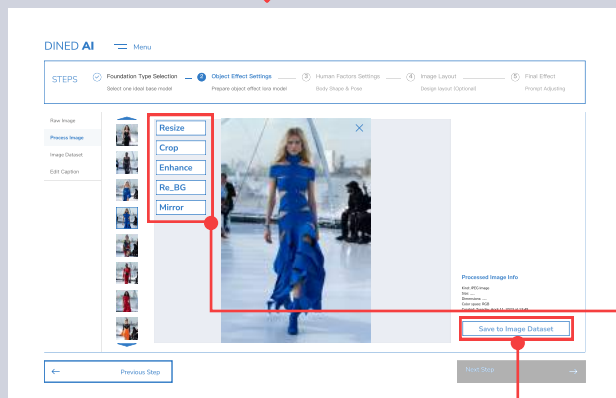


Figure 6.2.1-2 Process Image

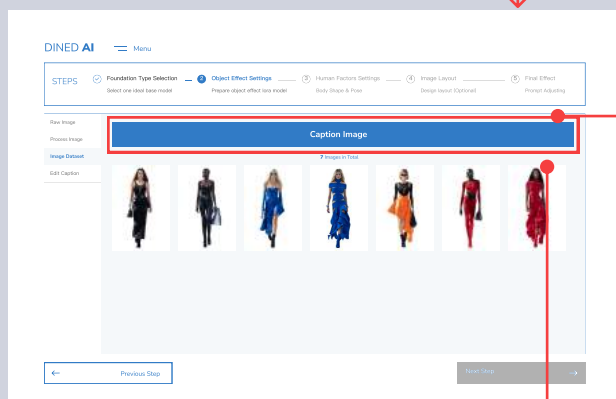


Figure 6.2.1-3 Image Dataset

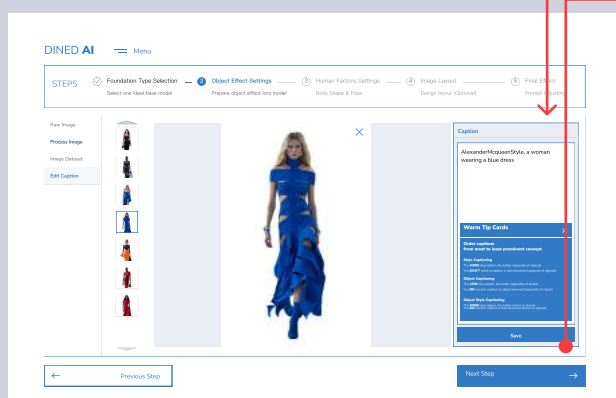
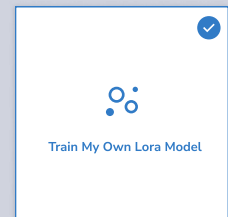


Figure 6.2.1-4 Edit Caption

S1: Upload Image

Images can be uploaded by dragging to the area and browsing local files



How to train your own lora model?

The next crucial step is configuring the Object Effect Settings. This particular stage empowers users to harness the capabilities of a loRA model, which acts as a guiding force for the stable diffusion model, resulting in the generation of images with specific directional attributes. Users have the flexibility to choose from a diverse range of pre-existing loRA models available online or train their own customized loRA model using the comprehensive resources offered by DINED AI.

S2: Five Functions to Process Images

Resize: Use length and width data to change the image size

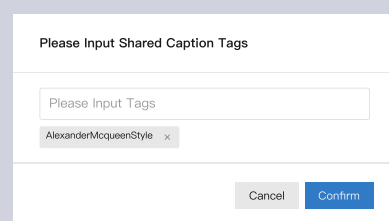
Crop: Manually crop the image size

Enhance: Use AI to enhance the image quality

Re_BG: Remove background

Mirror: It can be mirrored horizontally and vertically

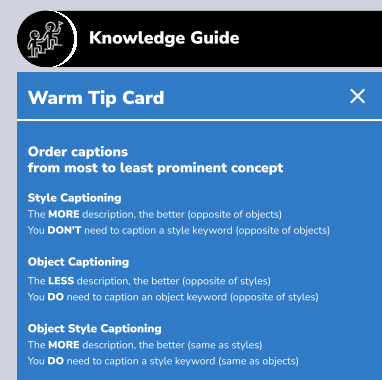
S3: Tag Images Automatically



Pop-up: Allow users to add unified tags or leave this option blank

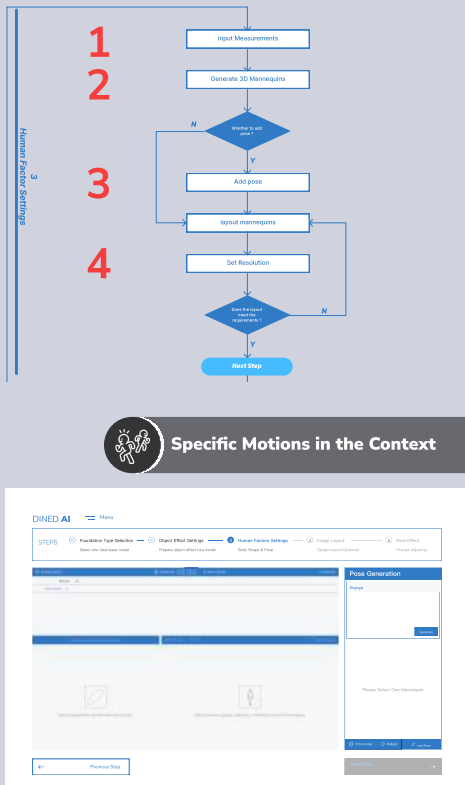
S4: Tag Images Manually

Following the **guidance** in this tool, you can manually adjust the labels corresponding to the pictures based on the automatically generated labels.



6.3 STEP_3_Human Factors Settings

User Flow



Step 3 : Human Factors Settings

The step serves as the central session that seamlessly connects anthropometry knowledge in DINED with SD's AIGC ability. This pivotal step empowers users to generate a wide range of multi-targeted body shapes that precisely match realistic statistics within their specific project context. Moreover, it offers users the capability to incorporate various poses into the body shape, marking a significant milestone in assisting designers with the dynamic design of human factors. By integrating anthropometry knowledge into DINED and leveraging SD's AIGC ability, this step enables users to explore an extensive spectrum of body shapes, ensuring a comprehensive and realistic representation. This feature is invaluable as it facilitates a deeper understanding of human proportions and dimensions, essential for designing products or environments that cater to the diverse needs and requirements of different individuals.

Design Point 3

1.The DINED product is directly embedded: The author embed the DINED original product interface directly into the new product, incorporating the functions provided by DINED into the new workflow. The benefit of this approach is that as DINED products are optimized, DINED AI becomes directly profitable.

2.Addition of text-to-pose function [1]: Building upon the existing functions of DINED, the author has introduced a new feature called text-to-pose to meet specific pose requirements in the design process. This design aspect can be considered a partial implementation of a particular motion design.

Reasons for Not Including Dynamic Motion Functions

The implementing dynamic motion functions would exponentially increase the tool's usability complexity. Pose generation can be a component of motion generation according to MDM method [2]. Initially, the pose capability can be developed in the Alpha version of the DINED AI product, while optimizing the user experience for motion generation in future development. The output content format (such as dynamic video files) does not align with the author's current AI capabilities for content output (text-to-image, picture files). However, motion output can be achieved through the text-to-image capability [3].

1: Input Measurements to Generate 3D Mannequin

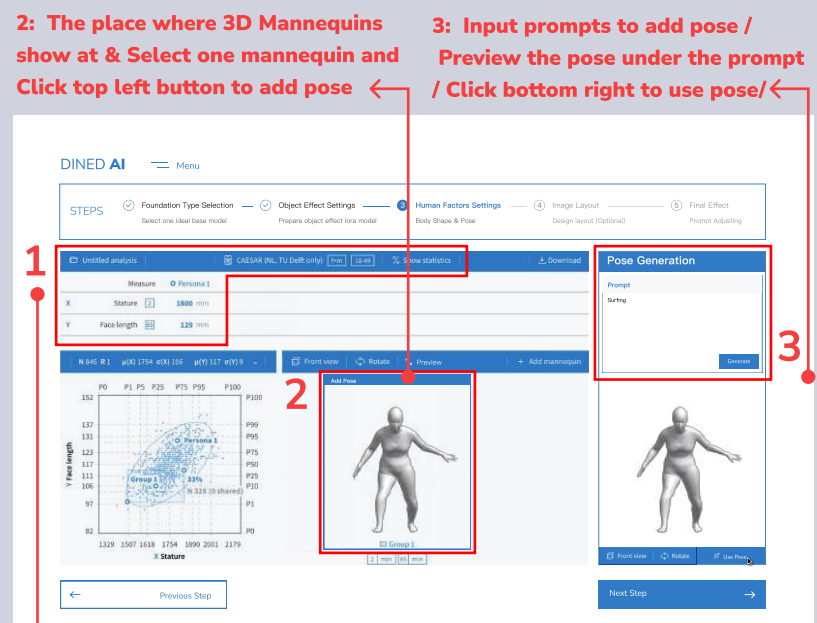


Figure 6.3-2 Situation adding pose to DINED Mannequin

6.3 Reference

- [1] Roy, P., Ghosh, S., Bhattacharya, S., Pal, U., & Blumenstein, M. (2022, October). TIPS: Text-Induced Pose Synthesis. In Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022. Proceedings, Part XXXVIII (pp. 161–178). Cham: Springer Nature Switzerland.
- [2] Tevet, G., Raab, S., Gordon, B., Shafir, Y., Cohen-Or, D., & Bertano, A. H. (2022). Human motion diffusion model. arXiv preprint arXiv:2209.14916.
- [3] Admin, & Admin. (2023, March 29). AIGC, is Quietly Changing the Short Video Industry. Cutout.pro. <https://www.cutout.pro/blog/2023/03/29/aigc-is-changing-short-video-industry>

4: Click Preview Button to Open a New Pop-up window with a Larger view of Mannequins

How to Change Perspective in Preview

Hover on one mannequin and you can see a control bar on the top; You can change the perspective of the mannequin. The mannequin can be also deleted on the tool bar.

Create a multi-view picture for one mannequin

A multi-view picture is an effective method utilized by designers to showcase their ideas. This technique allows for the display of different perspectives of a design on a single mannequin. In situations where only one persona mannequin is available, it is possible to incorporate additional mannequins. By utilizing the "Add The Same Mannequin" button, users can easily include more identical mannequins. The control bar provides the means to switch between various views, enabling users to create a comprehensive three-view picture of their design.

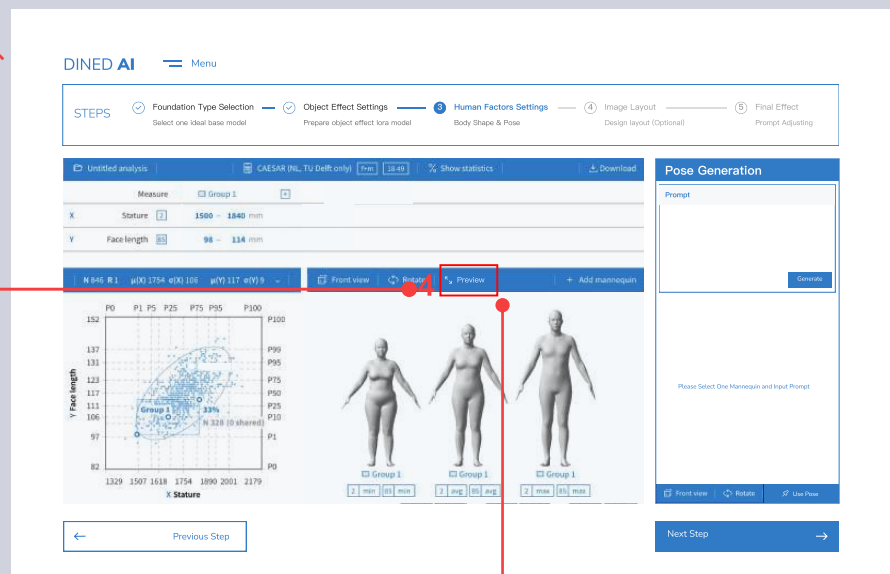


Figure 6.3-3 Situation adding multiple mannequins

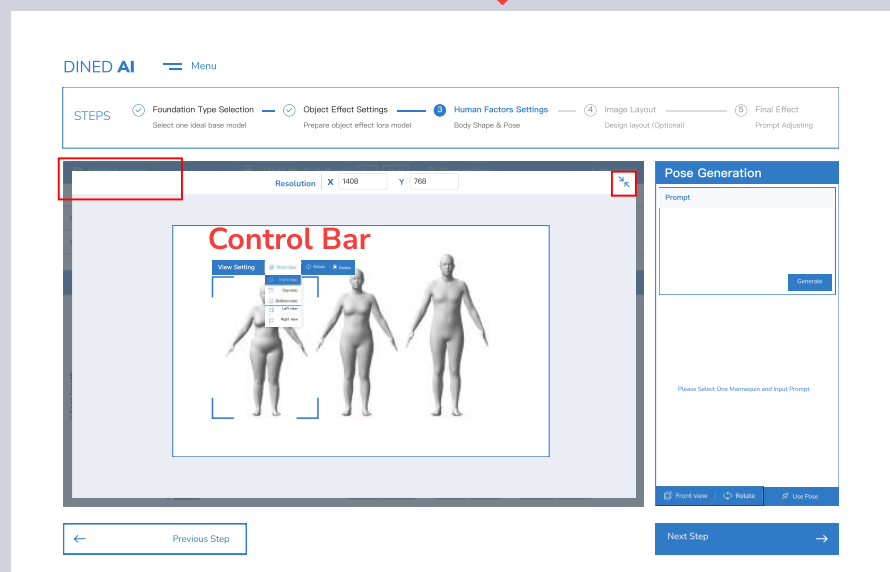


Figure 6.3-4 Preview of Group Only

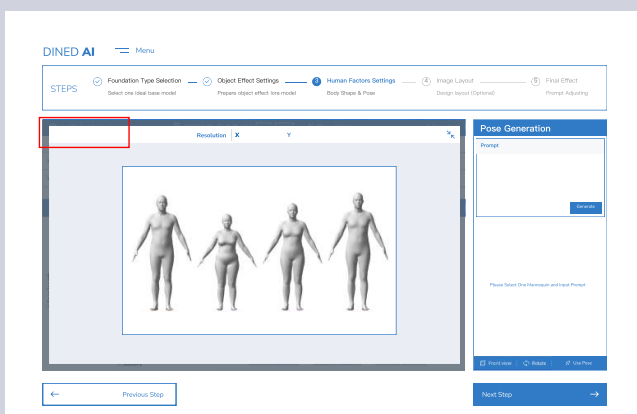


Figure 6.3-5 Preview of Group and Personas

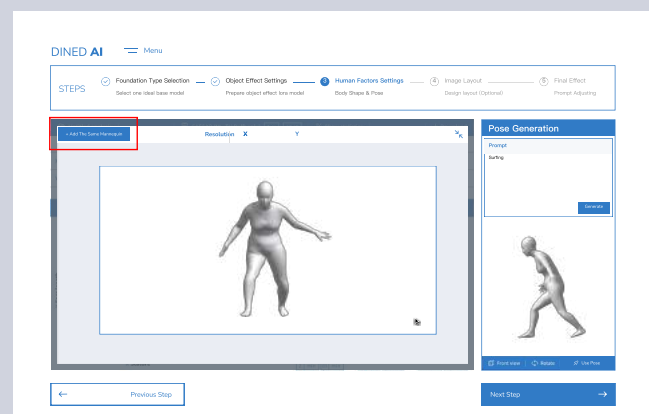
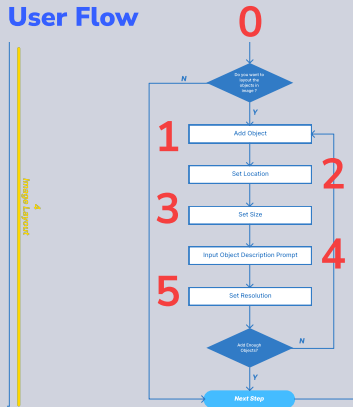


Figure 6.3-6 Preview of One Persona Only

6.4 STEP_4_ Image Layout (Optional)

User Flow



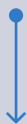
Step 4: Image Layout (Optional)

This step enables users to utilize color block and sub-division prompts for organizing the objects within the image. It's important to note that users have the option to skip this step, as it does not impact the visual outcome of the picture provided in step 5.

1. Add Object

S1: Add Objects

S2: Pick Color for Objects

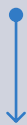


2. Set Location

3. Set Size

S3: Draw / Input Value to Create A Segmentation

S4: Tag the Segmentation



4. Set Resolution

Step 4 keeps the Resolution in Steps 3

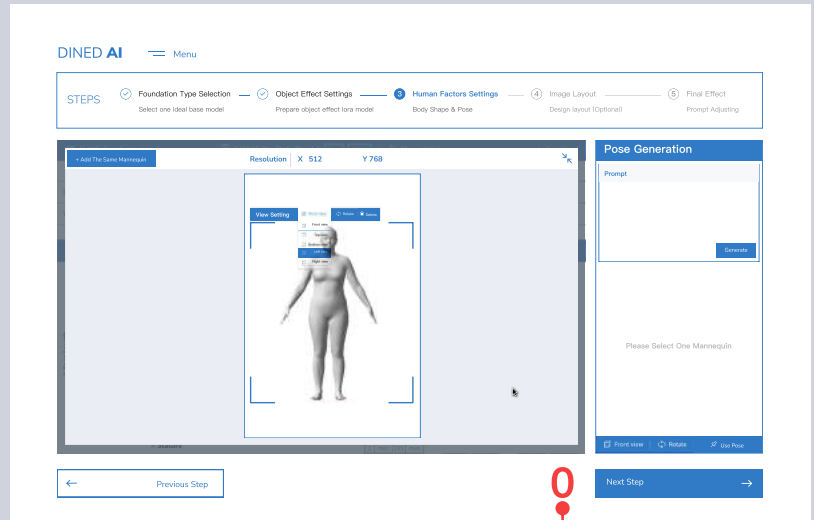


Figure 6.4-1 Step 4 keeps the Resolution in Steps 3

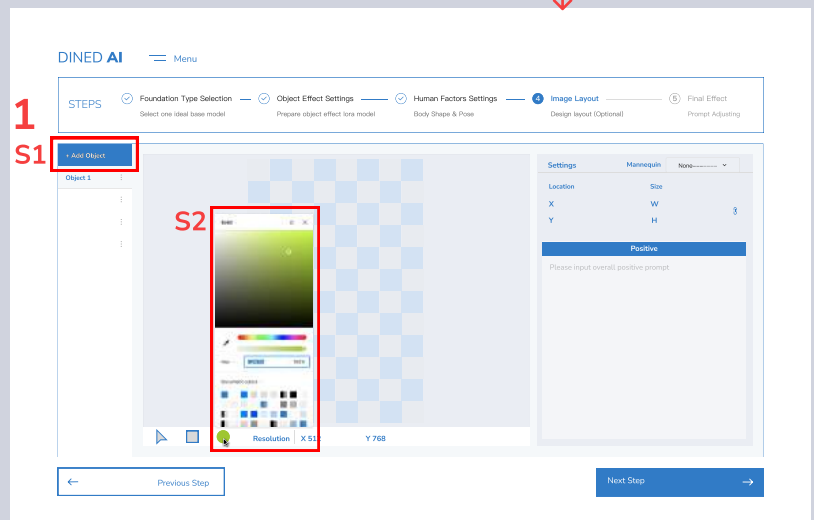


Figure 6.4-2 Select Color for the Segmentation

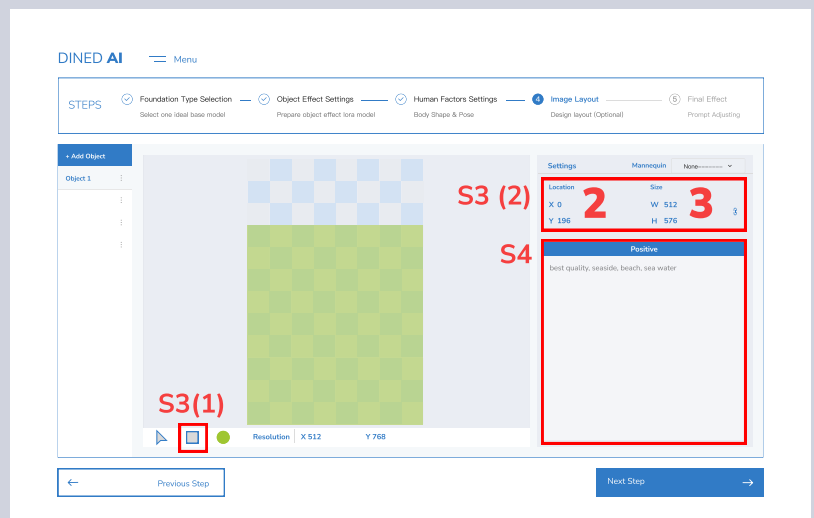


Figure 6.4-3 Drag or Input Value to Draw Segmentations

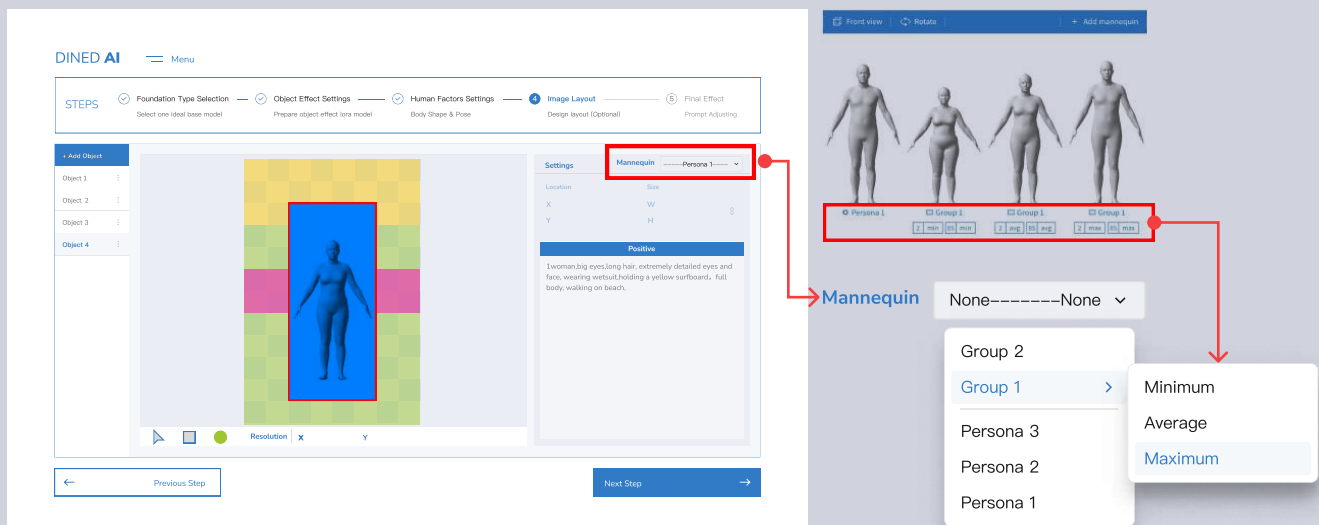


Figure 6.4-4 Mannequin Selection Component

To add a subdivision prompt to a mannequin, users can follow these steps:

1. Click on the drop-down menu located at the top right corner.
2. From the options available, select the desired mannequin name.
3. After selecting the mannequin, a sub-division prompt will be added automatically.
4. In this step, the input boxes for location and size values are disabled. This is because these properties of the mannequins are already set in the previous step.

By following these instructions, users can easily add a sub-division prompt to a specific mannequin while ensuring that the location and size values remain consistent with the previously set properties.

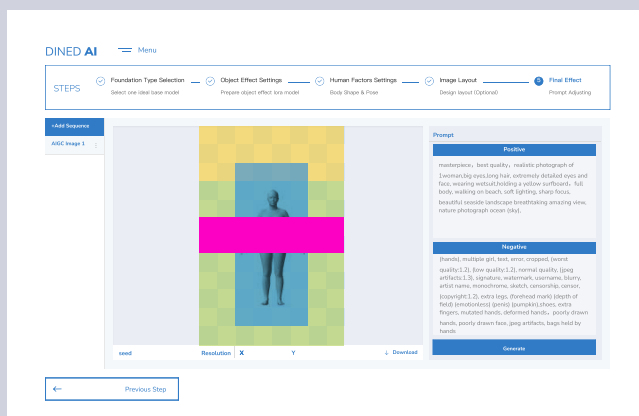


Figure 6.4-5 The interface when users enter the Step 5 after going through the Step 4

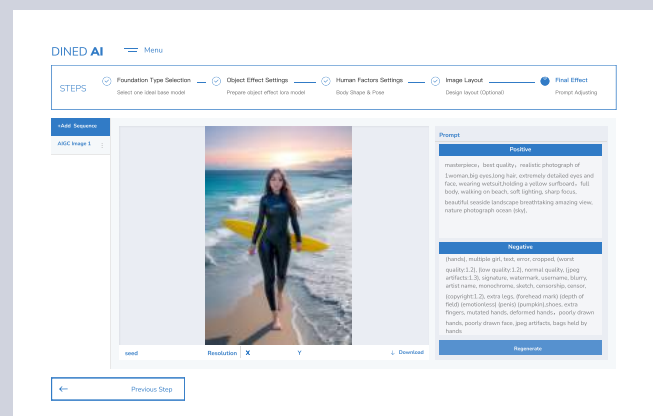


Figure 6.4-5 The image output in workflow including Step 4



Equipment

Design Point 4

Add MultiAreaConditioning Function: This approach enables the efficient arrangement of the necessary equipment within the given context. It provides designers with a convenient means to manage the intricate interplay between products, users, and the environment in inspiration stage. This particular design point can be regarded as fulfilling user requirements during the design process, taking into account the relevant equipment. However, there is still ample room for improvement when it comes to dimensional accuracy.

6.5 STEP_5_Final Effect

User Flow

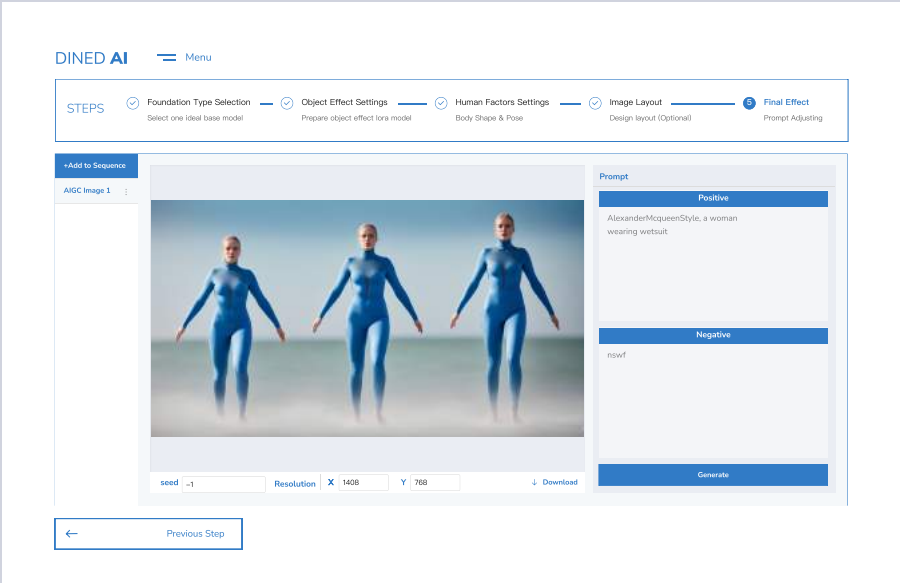
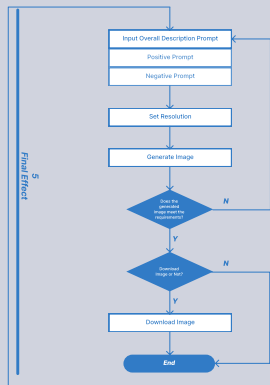


Figure 6.1-1 Home Page of DINED AI

Step 5 : Final Effect

The step enables users to generate images using prompts. This step is the closest to other SD GUIs.

How to Generate Image in Step 5 ?

With a straightforward process, users can easily harness the power of AI to generate personalized visuals. The following steps outline the path to creating remarkable AIGC images:

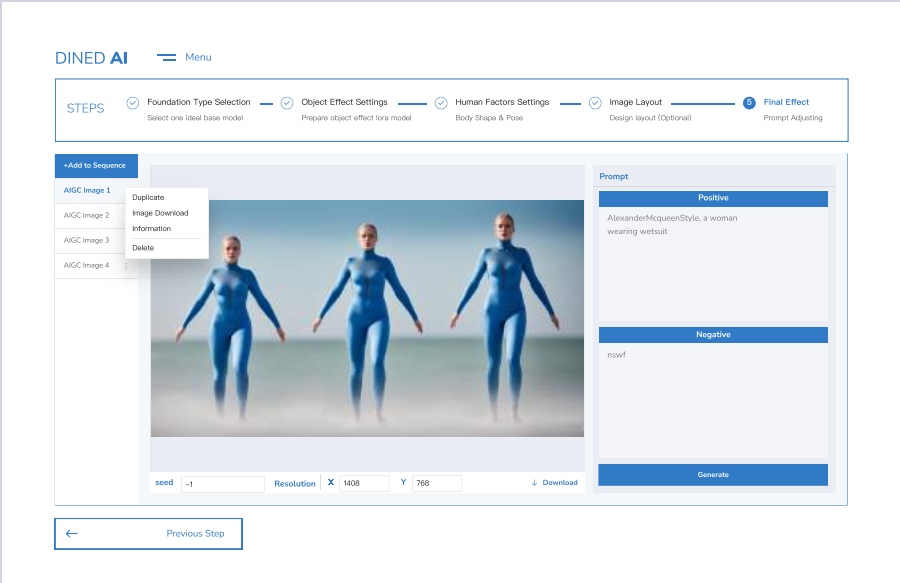


Figure 6.1-2 User Flow

Step5_Step 1:

Adding a Task to the Sequence To begin, users must add a task to the sequence. This task serves as the foundation for generating the desired image. By defining the parameters and objectives of the task, users can provide clear instructions for the AI system.

Step5_Step 2:

Inputting Positive and Negative Prompts Within the task, users are required to provide positive and negative prompts. These prompts play a crucial role in guiding the AI system's image generation process. By carefully crafting positive prompts that highlight the desired elements and negative prompts that specify what to avoid, users can effectively shape the outcome of the generated image.

By following these steps, users can tap into the remarkable capabilities of AI to create visually captivating and personalized AIGC images that meet their specific requirements.

5 Steps

STEP_1

Foundation Type Selection

STEP_2

Object Effect Settings

STEP_3

Human Factors Settings

STEP_4

Image Layout (Optional)

STEP_5

Final Effect

4 Design Points

Design Point 1 : Partner with major free and open source websites

Integrate with Civit AI

Design Point 2 : Optimize user training operation flow

Involve LoRA Model

Design Point 3: Add pose to pave ways to motion generation

1.The DINED product is directly embedded

2.Addition of text-to-pose function

Design Point 4: Manage the correlation between context elements and equipment

Add MultiAreaConditioning Function

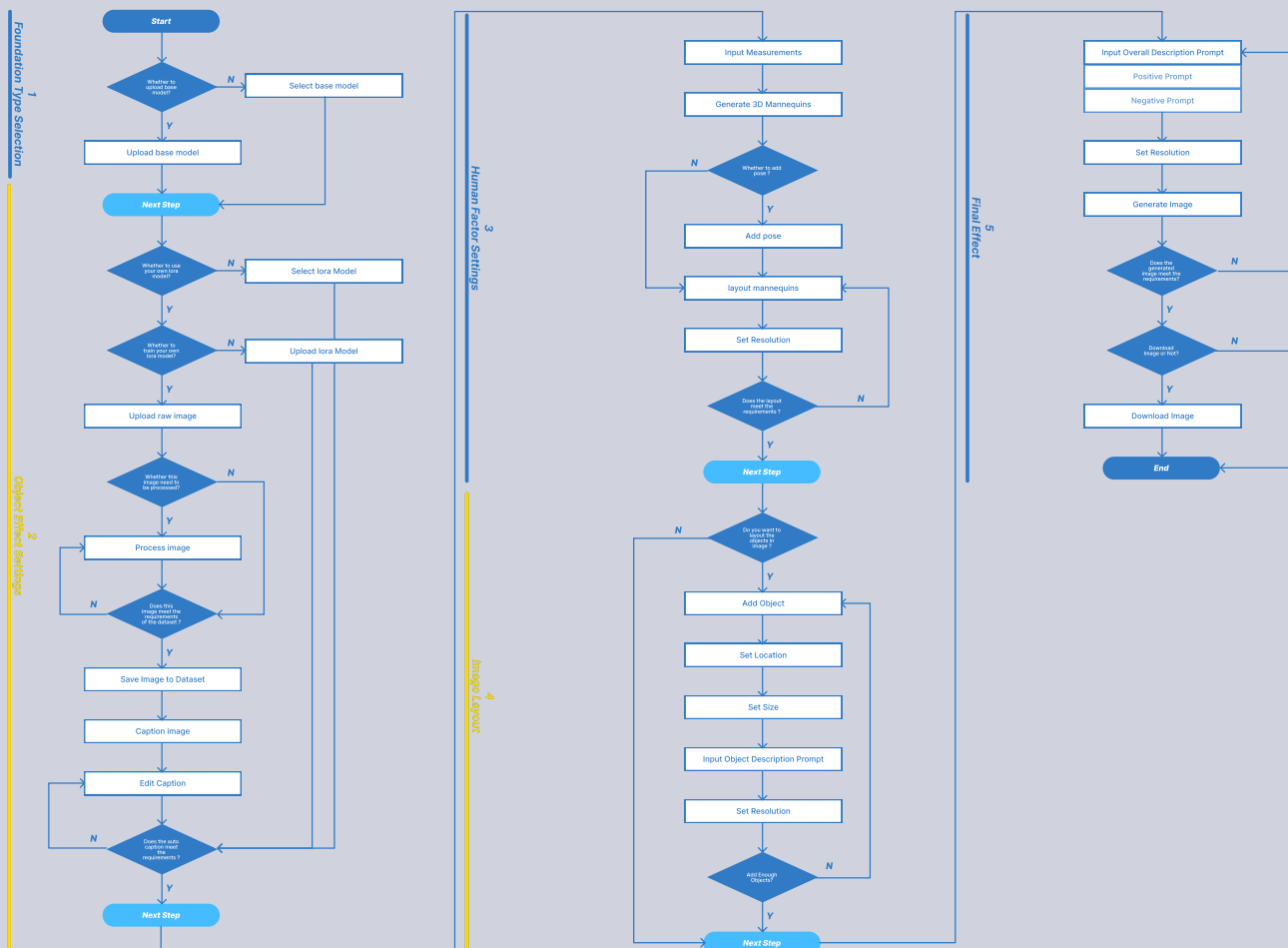


Figure 6.6-1 User Flow There are five steps in this DINED AI tool

07

Research Insights



Chapter 7 delves into a substantial body of research insights obtained through the AIGC process, providing a comprehensive exploration of their significance and implications.

Chapter Overview

7.1 Insights on ControlNet

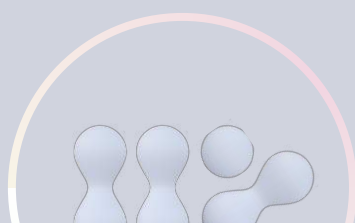
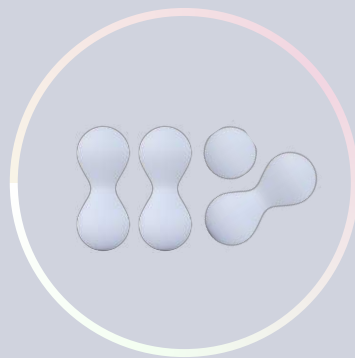
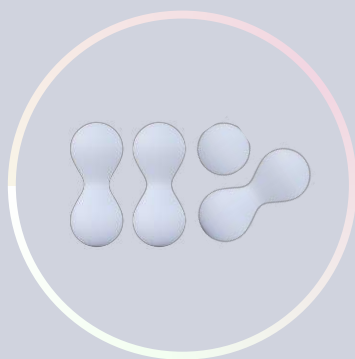
7.2 Insights on LoRA Model Training

7.3 Obese Body Shape & Base Model

7.4 Overall Comparison

Take Away

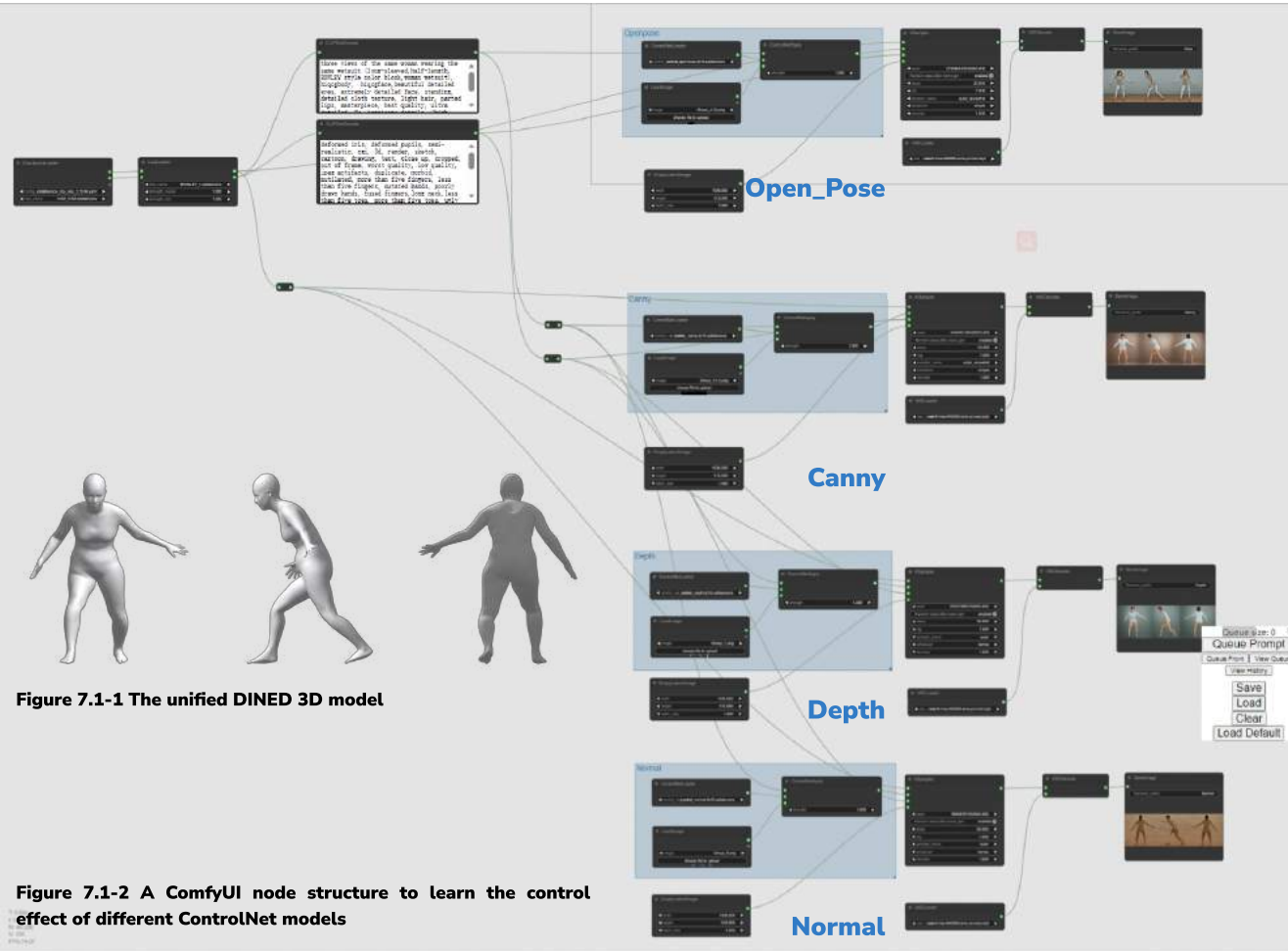
Reference



7.1 Insights on ControlNet

ControlNet is primarily utilized for body shape and posture control. In order to learn the control effect of different ControlNet models, a ComfyUI node structure was devised. In this analysis, I compared different models using the following approach:

- 1. Retaining the base model and prompts, we employ the unified DINED 3D model to generate a exported open_pose skeleton image, canny map, depth map, and normal map.
- 2. Four ControlNet control models are interconnected in parallel.
- 3. Simultaneously generate four models under ControlNet control. The outcome of this parallel connection is as follows:



This research exclusively focuses on comparing the algorithm's control quality regarding the generated character's body. Consequently, it does not encompass a detailed comparison of the character's face, feet, and hands, resulting in the absence of relevant optimization nodes. Moreover, due to the limited number of generated pixels, there may be certain challenges in accurately depicting the facial features of the experimental results. Therefore, the display results do not include any information related to face generation.

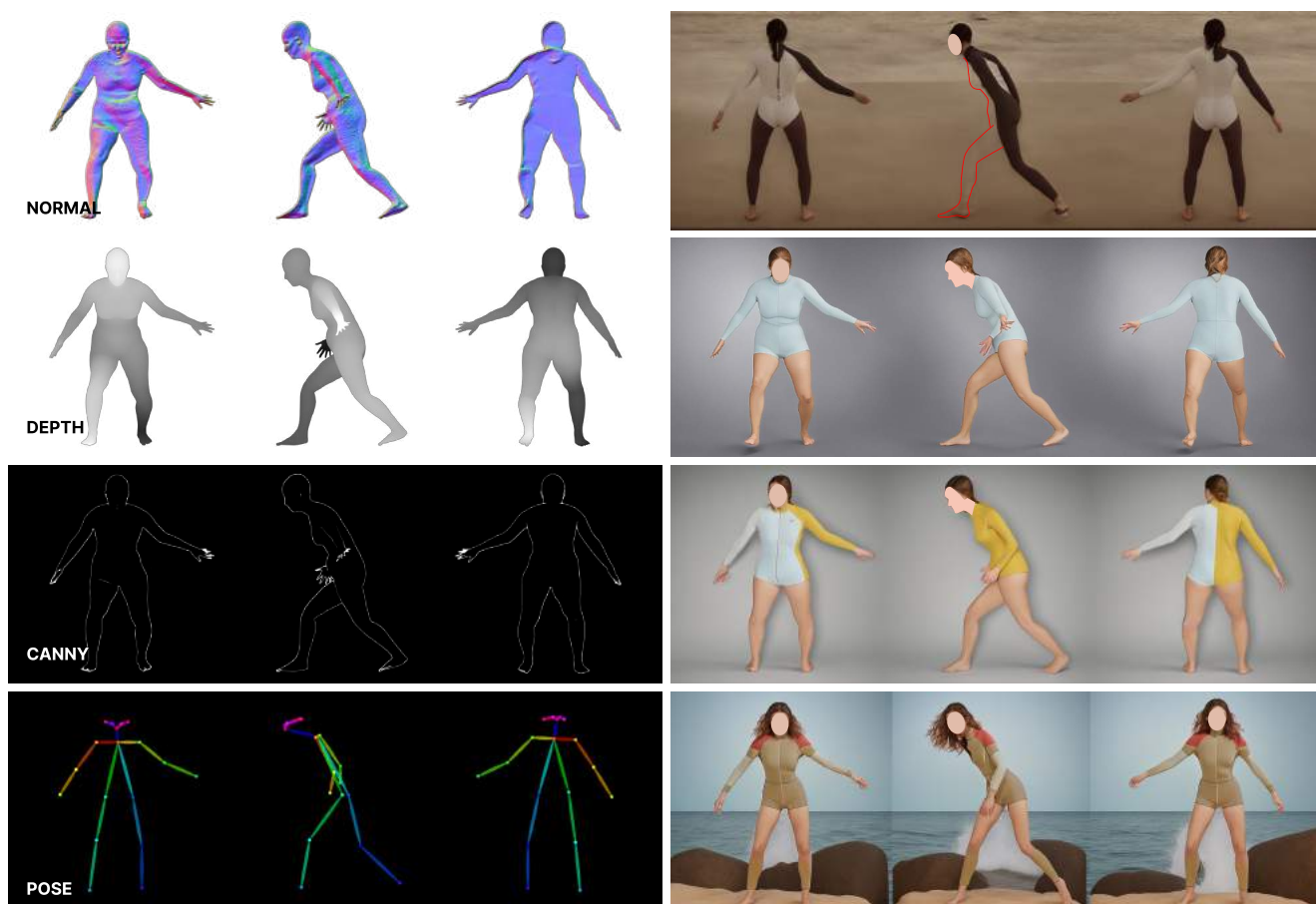


Figure 7.1-3 Output of Four models (more results please see appendix)

Observing the figure above, the following observations can be made:

1. Open pose generates images with better aesthetics. However, the pose of the character differs from the original DINED 3D model. And Open pose can't describe the body shape.
2. The Canny Model accurately replicates the pose of the original DINED 3D model but struggles to effectively distinguish the character's front and back position. In the front view, the front and rear positions of the two legs are not reflected in the generated picture.
3. The Depth Model exhibits satisfactory generation results, correctly recognizing gestures and distinguishing between the character's front and rear views.
4. The Normal Map technique produces the lowest quality results, struggling to accurately depict the body contours. Further investigation reveals that the issue lies in the export of the normal map. The author utilizes an online normal map generation tool, which may not meet the requirements of ControlNet.

Attempting alternative methods for generating the Normal Map:

1. Method 1-Employing Spline to generate a Normal Map still yields unsatisfactory results. Based on the generated outputs, it is speculated that there might be a disparity in the XYZ direction color settings between ControlNet and the generated Normal Map.
2. Method 2-Utilizing ControlNet's embedded method to generate normal map still fails to meet the required criteria. When using the output of Depth Model (see Figure 7.1-6) as input, the normal map generated is more three-dimensional than that generated when using screenshot of original DINED 3D mannequin as an input (see Figure 7.1-5). This discrepancy may arise from ControlNet's current normal map algorithm, which differs from generating a normal map from a 3D model.

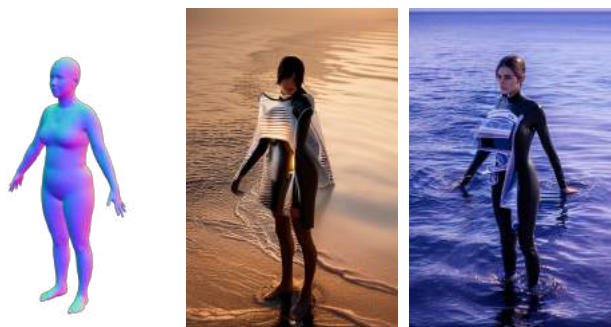


Figure 7.1-4 Method 1: The first image is a Normal Map generate from the original DINED 3D mannequin with Spline <https://app.spline.design/home> online tool. The results are still problematic.

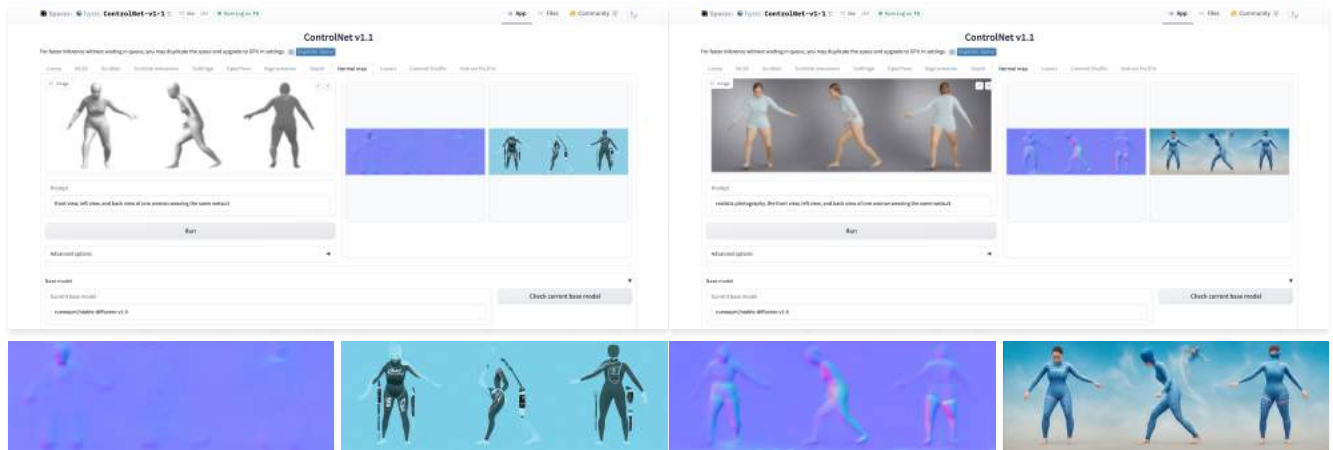


Figure 7.1-5 Method 2.1 Utilizing Screenshot of original DINED 3D Mannequin with ControlNet's embedded method to generate normal map

Figure 7.1-6 Method 2.2 Utilizing output image of Depth Model with ControlNet's embedded method to generate normal map

In conclusion, based on the aforementioned results,

1. The **Depth Model** exhibits the most favorable generation outcome, which not only able to depicts the body shape but also good at dealing with the front and rear positional relationship of posture.
2. The threshold for obtaining a normal map that meets the requirements of ControNet is relatively high. Consequently, for further utilization, the focus will shift away from the Normal Map as the research subject and instead prioritize the remaining three models which are **Open_Pose Model**, **Canny Model**, and **Depth Model**.

Relatively Optimal Model: Depth Model

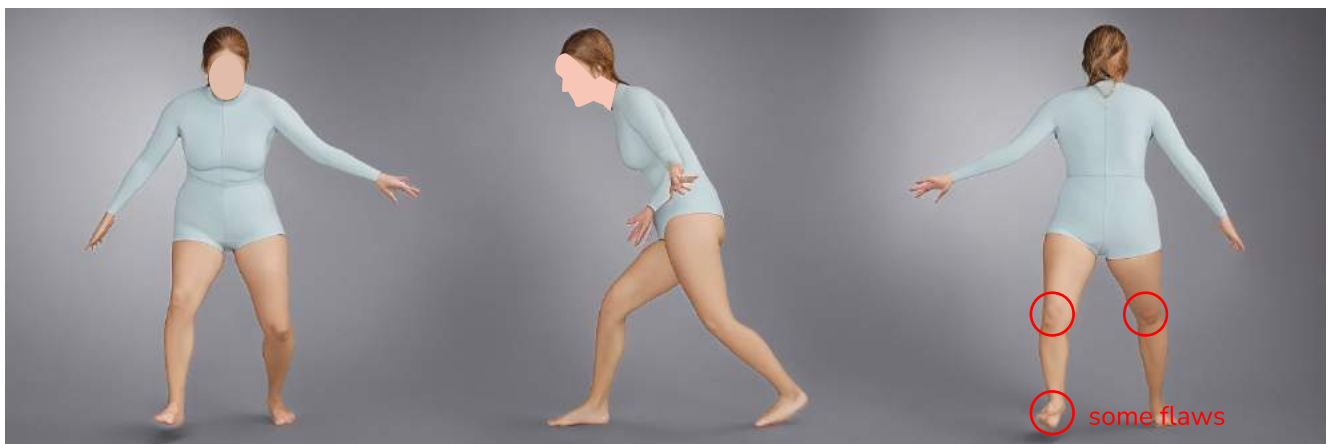


Figure 7.1-7 Flaws in Depth Model generation (Please See Appendix C7.1)

7.2 Insights on LoRa Model Training

As a trainer specializing in Kohya's Stable Diffusion, the author has gained valuable insights throughout the entire LoRA model training process.

The quality of the LoRA dataset is crucial for the performance of the LoRA model. It encompasses two main aspects: the quality of the dataset images and the quality of the dataset's picture annotations. The image quality of the dataset is determined by factors such as image resolution, pixel quality, and the total number of images. It is recommended that all images in the dataset are of the same size to ensure consistency. On the other hand, the quality of image annotation in the dataset is influenced by factors such as the level of detail provided in the image descriptions.

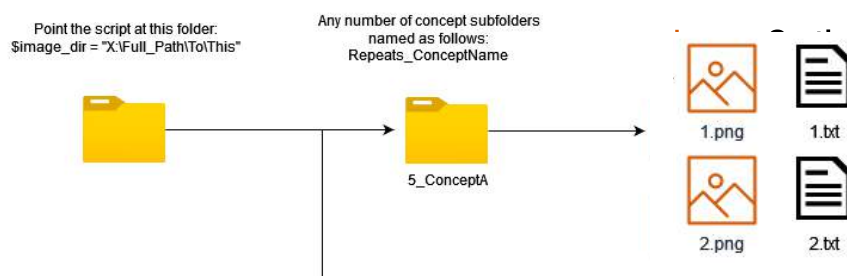


Figure 7.2-1 Image Directory Layout for training a LoRA Model [1]

Dataset Image

1. The images selected for the dataset should be carefully chosen to ensure that individuals can readily identify and categorize them under the same concept. This will facilitate easy testing of the reliability of the LoRA model's results.
2. It is advisable to preprocess the dataset photos by cleaning the background, removing unnecessary pixels, and reducing the workload during the subsequent captioning process. This step helps prevent the inadvertent labeling and contamination of the model throughout the captioning process.
3. In line with the desired generation outcome, the dataset should incorporate a significant number of materials with similar perspectives. For instance, if the objective is to generate portraits with side angles, the dataset should encompass a substantial collection of side-angle portraits to maximize the desired generation effect.

Caption

BLIP: Kohya provides an auto-tagging function called BLIP. BLIP (Better Language Model Image Prompts) is a method for generating captions for images using a pre-trained language model. It can be used in conjunction with Kohya's image captioning feature to generate captions for your images [2]. However, some users have reported issues with BLIP captioning, such as generating inaccurate or repetitive captions. As an alternative, some users prefer to manually create captions and save them in a text file [3]. **Different LoRA types require different caption methods (see Chapter 5, 5.4).** In this project, the author intends to train object style LoRA which means the picture annotations should be as comprehensive as possible, particularly highlighting elements that are not intended to be included in the generated result. By clearly marking such elements, it can be ensured that the unwanted elements do not appear in the generated output unless explicitly mentioned in the positive prompt.

Insights About Overfitting & Generalisation of the Model

If the dataset contains a limited number of images or if the images within the dataset are excessively similar, it may lead to overfitting of the LoRA model. In such cases, the generated image features will closely resemble those of the original dataset images. Conversely, if the dataset comprises an excessive number of images with a wide range of styles or if the annotations are ineffective or insufficiently detailed, it may hinder the convergence of LoRA training. Utilizing a LoRA model trained under these circumstances will result in generated pictures that fail to reflect the distinctive characteristics of the LoRA model.



Figure 7.2-2 Failed LoRA Model: Incorrect Material and Ill-Fitting Clothes in Generated Full-Body Image In this figure, we present an example of a failed implementation of the LoRA model. The left image, also generated via ComfyUI, attempts to replicate the appearance of actual clothing, while the middle images comprise the dataset used to train the LoRA model. On the right side, we observe the inaccurately sized garments in the generated full-body image, further highlighting the model's deficiencies.

Insights About Training Clothing & Characters

During the training process, the image is treated as a unified entity, with the clothing being an essential component of the character. The character's size is used as a point of reference for determining the suitable dimensions of the garments. When training the Clothing LoRA model, it is advisable to ensure that the clothing is properly worn on the character. This ensures that the resulting garment sizes are reasonable and practical. Neglecting to do so can result in inaccurate and impractical sizing outcomes for the garments.

7.2 Reference

- [1] 8.6 KB folder on MEGA. (n.d.). <https://mega.nz/folder/p5d3haJR#SmDSpaldBGcYzvZOx8sqbg>
- [2] Bmaltais. (n.d.-b). Unable to Get BLIP Captioning to Work · bmaltais/kohya_ss · Discussion #303. GitHub. https://github.com/bmaltais/kohya_ss/discussions/303
- [3] Kasuka. (n.d.). Reddit - Dive into anything. https://www.reddit.com/r/StableDiffusion/comments/118dq19/video_guide_for_training_lora_for_style_links_in/

7.3 Obese Body Shape & Base Model

Unintentional Bias in the Base Model: Due to its limited exposure to certain prompts, the base model appears to exhibit a lack of focus on fashion in relation to diverse body shapes, particularly when it comes to formal attire for individuals with fuller figures as opposed to those with slim body shapes. Consequently, when generating clothing options for individuals with fuller figures, the model tends to generate images that accentuate larger imperfections more frequently.

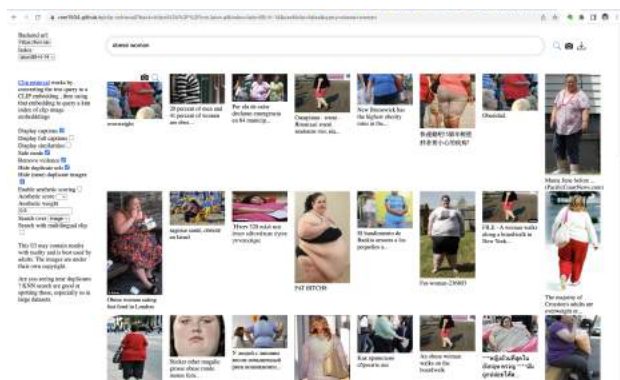


Figure 7.3-1 Search “Obese Woman” in LAION [1]

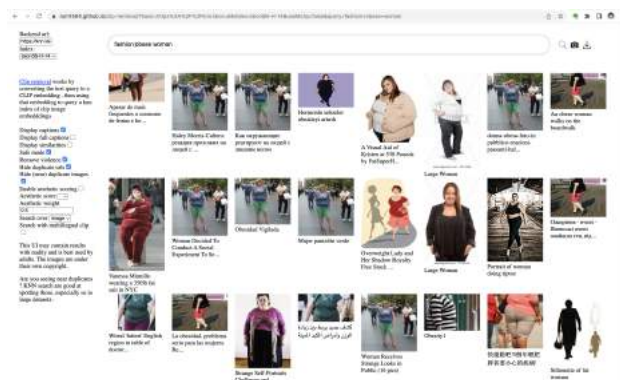


Figure 7.3-2 Search “Fashion Obese Woman” in LAION [2]

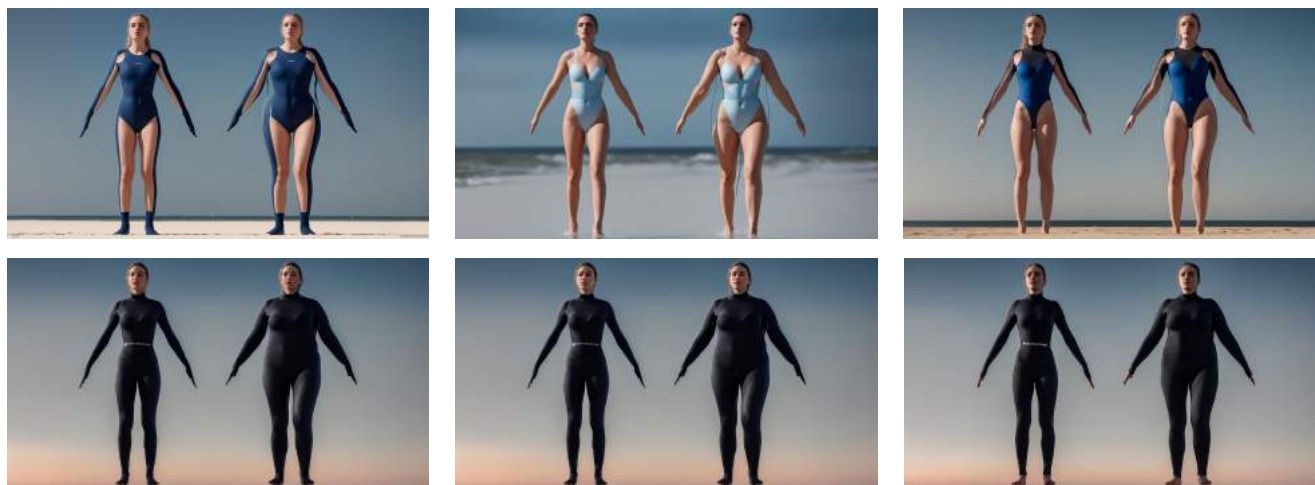


Figure 7.3-3 Unintentional bias Canny Results. The initial outcome from the first line reveals a design in clothing that exhibits significant variations. However, a flaw arises when attempting to generate an obese figure. In order to address this issue, the second line introduces an adjustment to the seed value. Consequently, the second attempt yields a result wherein the outline of the obese figure is accurately generated. Nevertheless, it becomes apparent that the clothing lacks distinct alterations or noteworthy changes in this iteration.

The large model itself possesses a particular flaw, which is further accentuated by the guidance provided through the use of LoRA material. This exacerbates the "prejudice" inherent in the model. However, it is worth noting that there is a scarcity of high-quality materials catering to overweight women seeking fashionable clothing. As a result, users can compensate for this limitation by adjusting the seed and continually refining their prompts, thereby mitigating the model's shortcomings.

7.3 Reference

- [1] Clip front. (n.d.). <https://rom1504.github.io/clip-retrieval/?back=https%3A%2F%2Fknn.laion.ai&index=laion5B-H-14&useMclip=false&query=obese+woman>
- [2] Clip front. (n.d.-b). <https://rom1504.github.io/clip-retrieval/?back=https%3A%2F%2Fknn.laion.ai&index=laion5B-H-14&useMclip=false&query=fashion+obese+woman>

7.4 Overall Comparison

In general (see Figure 7.4-1 Overall Comparison), the ControlNet_Depth algorithm has better results in generating body shape designs, the ControlNet_Canny algorithm has better results in generating clothing details.

Insights:

- [1]** For slim body shapes, maintaining a strength value of 1 produces good results. For obese body shapes, increasing the strength value is required to generate a shape that is closer to the input.
- [2]** The larger the strength value of ControlNet_Depth, the better the generated body shape. However, excessively large strength values can affect the depiction of clothing details.
- [3]** ControlNet_Canny algorithm is prone to leave the outlines of canny map. The optimal range for the strength of the ControlNet_Canny algorithm is smaller than that of ControlNet_Depth.
- [4]** The depth map extracted from the 3D model in this case does not contain background depth information. Therefore, the generated background using depth map and prompt will be relatively blurry. On the other hand, Canny is not affected by missing line information and can generate backgrounds relatively well.
- [5]** Based on the general base model SD1.5, the body shape generation from the front view of ControlNet is better than the body shape generation from the side image. The side view has more ambiguities in the perspective angle.

ControlNet(CN)

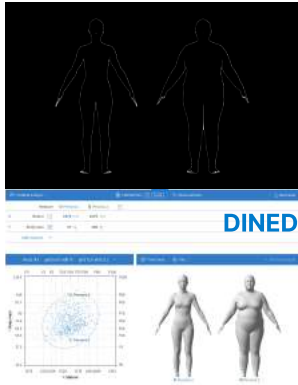
Canny

Input BodyShape

Prompts

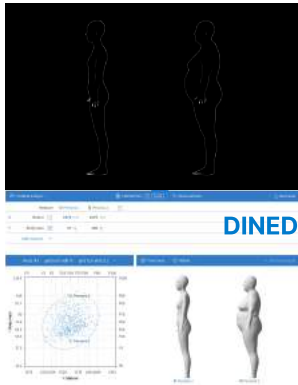
Canny_Front

3D Mannequin



Canny_Side

3D Mannequin



seed-82529106845132 (fixed) eluer_a / karras

Positive Prompt: realistic photography, 2 swedish girls, wearing exactly the same light blue AlexanderMcQueenSpring2023R2W style wetsuit, long-sleeved wetsuits, full-body wetsuits, full-length wetsuits, hiqcgbody, hiqcgface, beautiful detailed eyes, extremely detailed face, standing, detailed cloth texture, light hair, parted lips, looking at viewer, masterpiece, best quality, ultra detailed, 8k, intricate details, (high detailed skin:1.2), 8k uhd, dslr, soft lighting, high quality, film grain, dusk, seaside background, beach, front view, one obese woman on the right

Negative Prompt: deformed iris, deformed pupils, semi-realistic, cgi, 3d, render, sketch, cartoon, drawing, text, close up, cropped, out of frame, worst quality, low quality, jpeg artifacts, duplicate, morbid, mutilated, more than five fingers, less than five fingers, mutated hands, poorly drawn hands, fused fingers, long neck, less than five toes, more than five toes, ugly palms, ugly hands, ugly feet, poorly drawn face, cloned face, disfigured, mutation, deformed, blurry, dehydrated, bad anatomy, bad proportions, extra limbs, gross proportions, malformed limbs, missing arms, missing legs, extra arms, extra legs, weird colors, shoes, bag, jewellery, rings, bracelets, logo on wetsuit, text on wetsuit, bare legs, naked legs, different wetsuits

ControlNet Strength=1



seed-82529106845132 (fixed) eluer_a / karras

Positive Prompt: realistic photography, 2 swedish women, wearing exactly the same light blue AlexanderMcQueenSpring2023R2W style high neckline full-body wetsuit (long-sleeved, full-length, neoprene material, good seamlines, woman wetsuit), smooth swimsuit surface, hiqcgbody, hiqcgface, beautiful detailed eyes, extremely detailed face, standing, detailed cloth texture, light hair, parted lips, looking at viewer, masterpiece, best quality, ultra detailed, 8k, intricate details, (high detailed skin:1.2), 8k uhd, dslr, soft lighting, high quality, film grain, dusk, seaside background, beach, side view, standing,

Negative Prompt: deformed iris, deformed pupils, semi-realistic, cgi, 3d, render, sketch, cartoon, drawing, text, close up, cropped, out of frame, worst quality, low quality, jpeg artifacts, duplicate, morbid, mutilated, more than five fingers, less than five fingers, mutated hands, poorly drawn hands, fused fingers, long neck, less than five toes, more than five toes, ugly palms, ugly hands, ugly feet, poorly drawn face, cloned face, disfigured, mutation, deformed, blurry, dehydrated, bad anatomy, bad proportions, extra limbs, gross proportions, malformed limbs, missing arms, missing legs, extra arms, extra legs, weird colors, shoes, bag, jewellery, rings, bracelets, necklace, low neckline, logo on wetsuit, text on wetsuit, logo on whitesuit, wrinkles on wetsuit, ornament on the chest, different wetsuits, bared arms, bared legs, black outlines

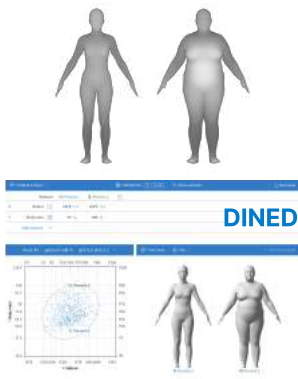
ControlNet Strength=1



Depth

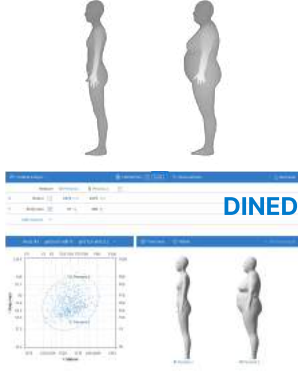
Depth_Front

3D Mannequin



Depth_Side

3D Mannequin



seed-82529106845132 (fixed) eluer_a / karras

Positive Prompt: realistic photography, 2 swedish women, wearing exactly the same light blue AlexanderMcQueenSpring2023R2W style high neckline wetsuit (long-sleeved, full-length, neoprene material, good seamlines, woman wetsuit), smooth swimsuit surface, hiqcgbody, hiqcgface, beautiful detailed eyes, extremely detailed face, standing, detailed cloth texture, light hair, parted lips, looking at viewer, masterpiece, best quality, ultra detailed, 8k, intricate details, (high detailed skin:1.2), 8k uhd, dslr, soft lighting, high quality, film grain, dusk, seaside background, beach, front view, standing,

Negative Prompt: deformed iris, deformed pupils, semi-realistic, cgi, 3d, render, sketch, cartoon, drawing, text, close up, cropped, out of frame, worst quality, low quality, jpeg artifacts, duplicate, morbid, mutilated, more than five fingers, less than five fingers, mutated hands, poorly drawn hands, fused fingers, long neck, less than five toes, more than five toes, ugly palms, ugly hands, ugly feet, poorly drawn face, cloned face, disfigured, mutation, deformed, blurry, dehydrated, bad anatomy, bad proportions, extra limbs, gross proportions, malformed limbs, missing arms, missing legs, extra arms, extra legs, weird colors, shoes, bag, jewellery, rings, bracelets, necklace, low neckline, logo on wetsuit, text on wetsuit, logo on whitesuit, wrinkles on wetsuit, exposed cleavage, hollow out on wetsuit, hole on front, ornament on the chest, different wetsuits

ControlNet Strength=1



seed-82529106845132 (fixed) eluer_a / karras

Positive Prompt: realistic photography, 2 swedish women, wearing exactly the same light blue AlexanderMcQueenSpring2023R2W style high neckline full-body wetsuit (long-sleeved, full-length, neoprene material, good seamlines, woman wetsuit), wetsuit with long sleeves, smooth swimsuit surface, hiqcgbody, hiqcgface, beautiful detailed eyes, extremely detailed face, standing, detailed cloth texture, light hair, parted lips, looking at viewer, masterpiece, best quality, ultra detailed, 8k, intricate details, (high detailed skin:1.2), 8k uhd, dslr, soft lighting, high quality, film grain, dusk, seaside background, beach, side view, standing,

Negative Prompt: deformed iris, deformed pupils, semi-realistic, cgi, 3d, render, sketch, cartoon, drawing, text, close up, cropped, out of frame, worst quality, low quality, jpeg artifacts, duplicate, morbid, mutilated, more than five fingers, less than five fingers, mutated hands, poorly drawn hands, fused fingers, long neck, less than five toes, more than five toes, ugly palms, ugly hands, ugly feet, poorly drawn face, cloned face, disfigured, mutation, deformed, blurry, dehydrated, bad anatomy, bad proportions, extra limbs, gross proportions, malformed limbs, missing arms, missing legs, extra arms, extra legs, weird colors, shoes, bag, jewellery, rings, bracelets, necklace, low neckline, logo on wetsuit, text on wetsuit, logo on whitesuit, wrinkles on wetsuit, ornament on the chest, different wetsuits, bared arms, bared legs,

ControlNet Strength=1



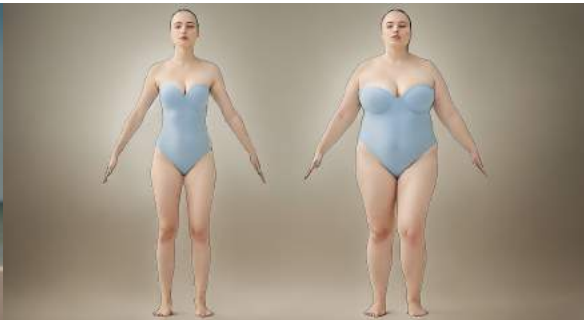
Figure 7.4-1 Overall Comparison.

Output Wetsuit Design

ControlNet Strength=2



ControlNet Strength=3



[3]

ControlNet Strength=2



ControlNet Strength=3

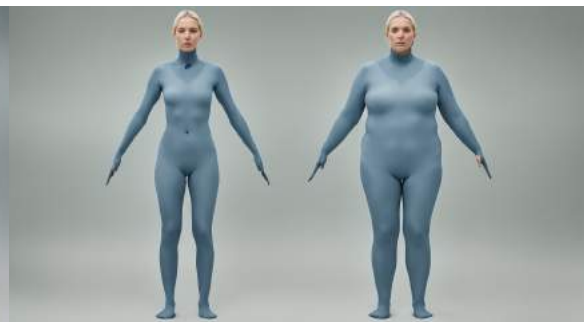


[1] [3]

ControlNet Strength=2



ControlNet Strength=3



[2] [4]

ControlNet Strength=2



ControlNet Strength=3



[1] [4]

Take Away

Summary 1: How different ControlNet models behave in this project

The Open_Pose Model improves image aesthetics but differs in character pose from the original DINED 3D model and does not describe body shape accurately.

The Canny Model replicates the pose accurately but struggles with front and back position distinction, particularly in leg positions.

The Depth Model produces satisfactory results by recognizing gestures and distinguishing between front and rear views.

The Normal Map technique generates lower quality results due to inaccurate depiction of body contours, potentially stemming from using an online tool for normal map generation that may not meet ControlNet requirements.

Summary 2: Advices for Training LoRA Model

Training images:

- Originals should be high-resolution, denoised, and then resized.
- Aim for a minimum of 10 images per concept, but more is highly recommended.
- Strive for maximum visual diversity while minimizing repetition, except for the object being trained.

Style captioning:

- Provide detailed descriptions for enhanced style understanding (opposite of objects).
- Arrange captions in order of prominence, from the most to the least prominent concept.
- Style keywords do not require captions (opposite of objects).
- Pay attention to specific word choice, as it is significant.

Object captioning:

- Keep descriptions concise for effective object representation (opposite of styles).
- Organize captions based on prominence, from the most to the least prominent concept (if applicable).
- Object keywords must be captioned (opposite of styles).
- Mindful selection of words is crucial for accurate representation.

Object style captioning:

- Provide comprehensive descriptions to convey object style effectively.
- Arrange captions in order of prominence, from the most to the least prominent concept (if applicable).
- Object keywords must be captioned (opposite of styles).
- Pay attention to specific word choice, as it plays a vital role.

Warm Tips for Training Clothing Model

During training, the image is treated as a whole, where clothing plays a vital role in character representation and sizing. The character's size serves as a reference for determining garment dimensions. Properly dressing the character in the Clothing LoRA model training is recommended to ensure reasonable and practical garment sizes. Neglecting this step may lead to inaccurate and impractical sizing results.

Reference (11)

Chapter 6

6.0 Reference

[1] TDesign. (n.d.-b). TDesign - Open source enterprise design system. <https://tdesign.tencent.com/index-en>

6.1 Reference

[1] Realistic Vision V3.0 - v3.0 (VAE) | Stable Diffusion Checkpoint | Civitai. (n.d.). <https://civitai.com/models/4201/realistic-vision-v30>

[2] Beautiful Realistic Asians - Brav5 | Stable Diffusion Checkpoint | Civitai. (n.d.). <https://civitai.com/models/25494/beautiful-realistic-asians>

6.3 Reference

[1] Roy, P., Ghosh, S., Bhattacharya, S., Pal, U., & Blumenstein, M. (2022, October). TIPS: Text-Induced Pose Synthesis. In Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXVIII (pp. 161–178). Cham: Springer Nature Switzerland.

[2] Tevet, G., Raab, S., Gordon, B., Shafir, Y., Cohen-Or, D., & Bermano, A. H. (2022). Human motion diffusion model. arXiv preprint arXiv:2209.14916.

[3] Admin, & Admin. (2023, March 29). AIGC, is Quietly Changing the Short Video Industry. Cutout.pro. <https://www.cutout.pro/blog/2023/03/29/aigc-is-changing-short-video-industry>

Chapter 7

7.2 Reference

[1] 8.6 KB folder on MEGA. (n.d.). <https://mega.nz/folder/p5d3haJR#SmDSpaldBGcYzvZOx8sqbg>

[2] Bmaltais. (n.d.-b). Unable to Get BLIP Captioning to Work · bmaltais/kohya_ss · Discussion #303. GitHub. https://github.com/bmaltais/kohya_ss/discussions/303

[3] Kasuka. (n.d.). Reddit - Dive into anything. https://www.reddit.com/r/StableDiffusion/comments/118dq19/video_guide_for_training_lora_for_style_links_in/

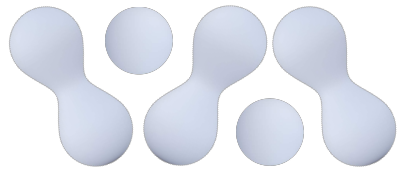
7.3 Reference

[1] Clip front. (n.d.). <https://rom1504.github.io/clip-retrieval/?back=https%3A%2F%2Fknn.laion.ai&index=laion5B-H-14&useMclip=false&query=obese+woman>

[2] Clip front. (n.d.-b). <https://rom1504.github.io/clip-retrieval/?back=https%3A%2F%2Fknn.laion.ai&index=laion5B-H-14&useMclip=false&query=fashion+obese+woman>

08

Evaluation



Chapter 8 aims to communicate the value proposition of the product designed to potential stakeholders, investors, and early adopters and illustrate how this technology can revolutionize certain aspects of the industry or solve existing challenges in innovative ways through expert interviews.

Chapter Overview

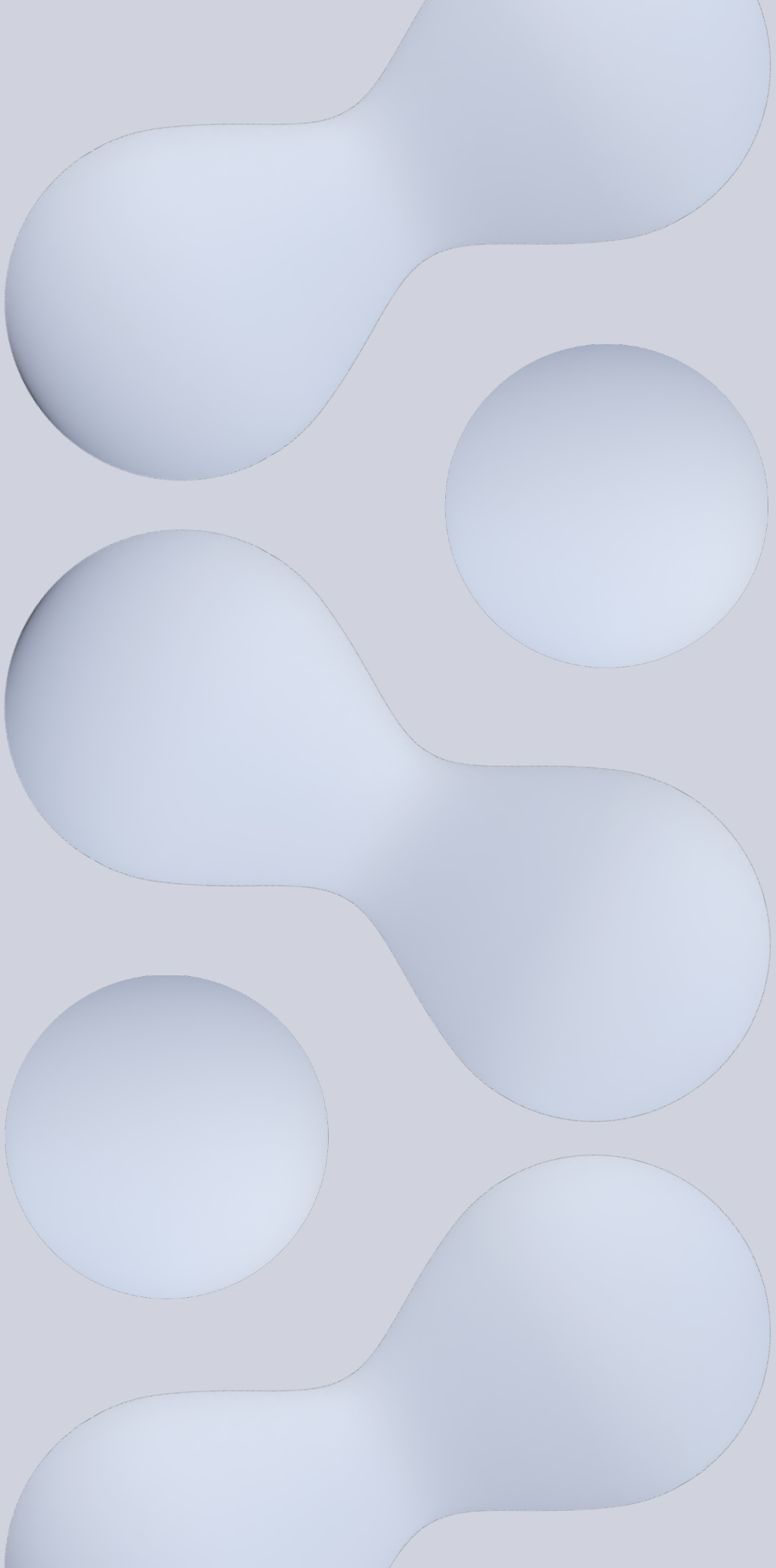
8.1 Evaluating With Experts

8.2 Expert Interview Brief 1

8.3 Expert Interview Brief 2

Conclusion

Take Away



8.1 Evaluating With Experts

Two expert interviews were hosted by the author to evaluate if the design is helpful for advanced designers. This series of expert interviews are devised to explore the “insider knowledge” of the fashion design expert and the wetsuit design expert, which can provide a deeper understanding of diverse body shape aesthetic related fashion design and soft sport product design of structural contexts.

Method: Open-end Structure Interview

Immersion in specialized knowledge and insights, experts are able to inform some knowledge that is not readily available in public sources, which makes preparing interview questions challenging. Actually, compared with structure interview, open-end structure interview allow researchers to adapt the conversation in real-time, enabling them to ask follow-up questions or delve deeper into specific topics.

Interview Cue

Though it is an open-end Structure Interview, there are three core topics marked as cues:

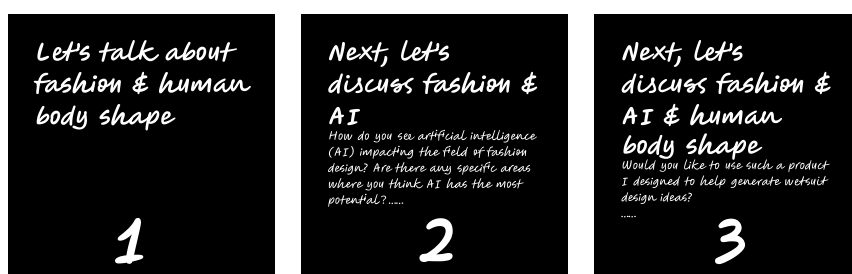


Figure 8.1-1 Interview Cue (Please See Appendix C8.1)

Alternative Question concerning Fashion and AI tool :

1. How do you see artificial intelligence (AI) impacting the field of fashion/ wetsuit design? Are there any specific areas where you think AI has the most potential?
2. In your opinion, what are some challenges or limitations that fashion/ wetsuit designers face today, and how do you think AI can help overcome them?
3. Could you provide some examples of how AI is currently being used in fashion/ wetsuit design? What are the key benefits or outcomes of incorporating AI into the design process?
4. How does AI influence the exploration and ideation phases of design? Are there any AI tools or technologies that you find particularly useful or innovative in this context?
5. With the rise of AI-generated clothing designs and virtual fashion models, what impact do you think this will have on the traditional role of designers? How can designers adapt to this changing landscape?

6. Are there any ethical considerations or potential drawbacks associated with the use of AI in fashion design? How can we ensure responsible and inclusive practices when integrating AI into the industry?
7. How can AI assist designers in improving sustainability and reducing the environmental impact of the fashion industry? Are there any specific AI applications or strategies you find promising in this regard?
8. Collaboration between humans and AI is becoming increasingly common in various fields. How do you envision the future of collaboration between designers and AI? What are the opportunities and challenges that may arise?
9. Are there any skills or knowledge areas that you believe designers should develop in order to effectively utilize AI in their work? How can designers prepare themselves for this evolving landscape?
10. As an expert in design, what advice would you give to aspiring designers who want to embrace AI as a tool in their creative process?



Figure 8.1-2 Inspiring Poster_Fashion Design

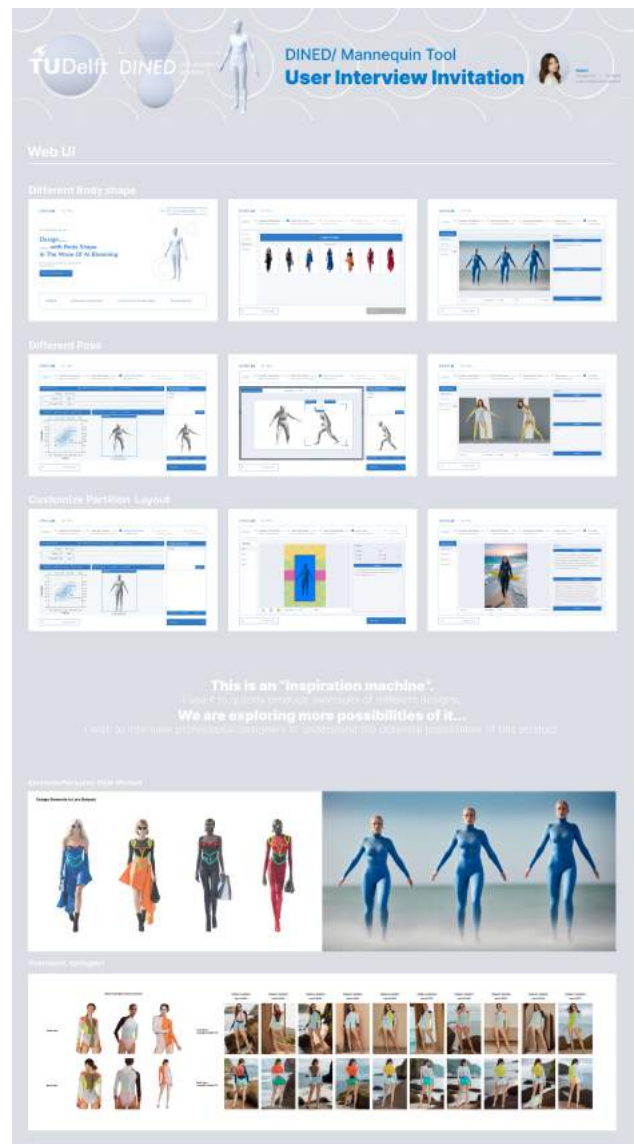


Figure 8.1-3 Inspiring Poster_Wetsuit Design

8.2 Expert Interview Brief 1

Dr. McQuillan, H.L. (Design Research Expert)

Academic Background:

Her research is in the field of multimorphic textile systems and encompasses the design and fabrication of animated textiles, 2D to 3D woven structures, design for circular systems and design methods. For over 15 years Holly has been investigating methods of ameliorating the negative impacts of the fashion and textile industry through holistic design and system practices.



- 1. Meet current market demand**
- 2. Have the potential to influence the iterative production and sales model**

Insights about Design Tool and Market business models

“SHEIN Inc. will like your project.”

SHEIN is an online fashion company renowned for its stylish apparel and fashion goods, predominantly marketed under the flagship "SHEIN" brand. Adopting a Direct Sales to Consumer (DTC) approach, SHEIN effectively eliminates the price spread by directly connecting with its customers. This streamlined process enables the company to swiftly incorporate customer feedback into its supply-demand dynamics, establishing a remarkably agile supply chain. [1]

Moreover, SHEIN leverages network technology to embrace a digital and modular fashion design model. As a prominent Chinese cross-border e-commerce brand, SHEIN utilizes real-time data collection to stay attuned to emerging fashion trends and competitive products. By establishing a design assistance system for its designers, SHEIN enables modular design within predefined standards and scope, thereby enhancing creativity and efficiency. This approach empowers designers to produce innovative designs promptly. Key Opinion Leader (KOL) marketing serves as a pivotal component of SHEIN's comprehensive marketing strategy. Collaborating with prominent social media platforms such as Facebook, Twitter, and Instagram, SHEIN leverages KOLs to amplify its brand presence. Additionally, the company capitalizes on various promotional opportunities such as pop-up store sales, KOL events, and fashion weeks. This multi-faceted approach helps SHEIN expand its reach and engage with its target audience effectively. [1]

Given SHEIN's business model, the incorporation of the DINED AI's inspiration generation aligns seamlessly with its objectives and operations. The DINED AI's capacity to generate innovative ideas perfectly complements SHEIN's agile and adaptable nature, further reinforcing its ability to meet the evolving demands of the fashion industry.

“ SHEIN put out hundred design idea and sketch options in its Sina Channel. Selected fans are invited as “taster”, to select which one is best and which one should be never got this kind. Shein has some stuff sentences following the sketch.

From Holly McQuillan

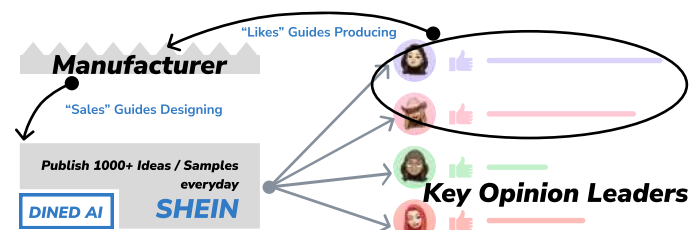


Figure 8.2-1 Successful Business Model in SHEIN and The Point DINED AI can be involved in

8.2 Reference

[1] Decoder. (n.d.). How did SHEIN become the top fashion seller of China's cross-border e-commerce? [www.linkedin.com. https://www.linkedin.com/pulse/how-did-shein-become-top-fashion-seller-chinas-cross-border-](https://www.linkedin.com/pulse/how-did-shein-become-top-fashion-seller-chinas-cross-border-)

“Such design might influence manufacture in return ”

"Let the clothes adapt to the people, rather than forcing people to conform to the clothes." is a customer-first approach that McQuillan believes can revolutionize the fashion industry. This innovative product has the potential to create a new kind of supply-demand market that is focused on inclusivity and diversity in sizing, rather than a limited number of standard sizes. This technology enables customers to not only try on clothes but also make adjustments to the fit of a model before the garments are produced in a factory. This has the potential to greatly reduce waste caused by ill-fitting clothing and empower customers with more control over their fashion choices. As this technology continues to develop, it will be exciting to see how the industry adapts to this more customer-centric approach.

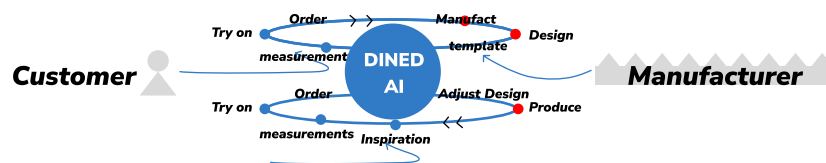


Figure 8.2-2 Potential Business Model with DINED AI

the blue points represent the stage can be achieved by DINED AI / the red points represent the stage charged by the manufacturer



1. Suitable for all ordinary users VS considering specially support current design habits of professional designers ?

Insights about the Target Group and User Experience

“Even for professional designers, in early design process, there are only vague definitions or words to describe the design.”

For figurative creation, prompts can be really effective in generating inspirations in early design stage. However, if the fashion designers are exploring abstract concept unknown, it is hard to guarantee the quality of AI generating images. In current design process, designers have to do lots of sketching to execute creating experiments on canvas. Nevertheless, in a inspiring tool like I designed, the experiment process for professional designers seems to be overlooked. Though, users can adjust output images via prompts and other controls, limited by fuzzy-front problem, the prompt description is lack of accuracy compared with sketching. For none-design background users, this problem will not influence the generating results too much. Because they have not been trained with thinking through sketching. While creating with nature language is just

“Sketching is currently a key aid to designers’ creating. Direct interaction with the design output will be more in line with the designer’s habits and intuition.”

From Holly McQuillan

what they are more familiar with. But for professional designers who have got accustomed to sketching thinking, they might find it hard to articulate the nuance design in their minds. In other words, if this tool consider professional designers into account, then DINED AI tool requires truly intuitive control in professional designer methods.

Are you willing to use the product I designed in your work?

“Fun to use it ! ”

but not necessary, for that I focus on fashion research rather than fashion design"

8.3 Expert Interview Brief 2

Ms. Timon Staal (Design Practice Expert)

Background:

DINED Product Early Rising Researcher

Who First Used DINED To Aid In 3D Body-fitting Design of Wetsuits

Wetsuit Technical Designer / Expert in Soft Sport Product Technical Designer



“I've tried stable diffusion as well. Um, just to generate a different results. But I didn't really get it to work so well maybe because I didn't have the time to look for how I can customize a model or or get it to work functionally for me. So the only things that I tried was just generate wetsuits in stable diffusion and Dall-E. And ChatGPT we've already some people using it here already for like our website and stuff. But for wetsuits, not yet.”

AI Tool Experience, From Timon Staal

Wetsuit Design Process

1. Designers start by working with existing wetsuits
2. The design process involves material design, pattern design, and graphic design
3. Material choices are based on stretchiness, thermal properties, price, and feel
4. Thickness varies depending on the part of the body for insulation and flexibility
5. Pattern design considers functionality and aesthetics
6. Some panels are necessary for closure and comfort
7. The overall design is similar to clothing design with considerations for shape and style



- 1. Super applicable for small design team**
- 2. Able to assist with project plan presentation and sampling process**

Insights about the Target Group

“Do you see generating inspiration part is important through the whole design process in your work?”

“Definitely. I must say we don't do it enough. We don't do a lot, because it is time consuming, there's a lot of pressure, we need a new design quickly, So sometimes we do it too rushed.”

From Staal Timon

Staal Timon is a member of a small design team, consisting of fewer than seven members. In this context, a small design team typically includes around three junior designers or even more. Despite its limited size, this team is entrusted with the responsibility of handling all the design tasks within the company.

Due to the limitation of project scale and the richness of design experience, such a design team has to learn from profession's design virtue.

It is an apparent short-circuit to learn the market trend and find their own brand definition.

Furthermore, the small team where Staal Timon works faces intense time pressures during the release of a new wetsuit product. The workload becomes overwhelmingly demanding, leaving them with limited time for each design process. As a result, the team often finds themselves delivering new designs hastily, relying on only a few design proposals during the initial stages. Despite Timon's understanding that following a formal design process yields superior results - one that involves generating a multitude of diverse ideas, engaging in collaborative discussions, and ultimately selecting the optimal direction - the constraints of time make it challenging to adhere to this approach.

When contemplating the provision of services to a small design team, it is essential for the DINED AI to be equipped with functionalities that facilitate seamless collaboration within the team. The integration of such features is paramount, as teamwork can make a profound difference in the realm of interaction design. Moreover, it is crucial to acknowledge that product complexity will invariably rise in tandem with effective collaboration.

Insights about Design Process:

Demonstration Before Project Approval

“It would be very helpful when you're trying to convince your superiors with a product idea, and it's really nice to generate an image like ‘OK! This is the product vibe tune.’ ”

While Timon Staal's company may not require slide presentations for higher-ups, numerous companies often demand designers to swiftly create visually captivating effects. This places an immense challenge on designers to produce renderings that accurately capture the desired product atmosphere even before the design project commences. When designers lack specific design guidance, this tool outperforms Photoshop in terms of speedily generating visual effects, providing a distinct advantage.

Sampling Process (Prototyping)

To compensate the missing of idea reflection, the team tweaks the design details in sampling process, which happens when the sample wetsuit are made. The design team will carefully compare every detail, and the members will quickly generate a series of alternatives and make quick modifications for the deficiencies.

“

But it (generate new design ideas) happens also a lot of times that we're during the sampling process, we still. Uh, you know you are. Wetsuit still changing a bit .But it happens also a lot of times that we're during the sampling process, we still. Uh, you know you are. Wetsuit still changing a bit .

”

From Staal Timon



- 1. Not supportive for wetsuit design details generation**
- 2. Can hardly convey the sport state of material's elasticity, thickness, and fit.....**

Insights about User Experience: Segmentation Model

There is insufficient support for the detailed design of the clothing. Timon's current role mainly involves overseeing the professional aspects of the wetsuit, such as modifying the anti-slip texture, optimizing the fabric technology, and balancing production costs, while staying within the company's established product style. These design changes may not significantly alter the appearance of the wetsuit. However, considering the current limitations of available tools, accurately depicting these intricate details in visual representations is challenging.

Due to the aforementioned reasons, Timon desires a feature that would allow for the marking of specific elements in the design of the wetsuit. For example, he envisions using a long blue rectangle to represent the zipper, while a color block on the side could indicate the presence of a non-slip texture.

To achieve this effect, an AI model known as a segmentation model would be ideal. However, incorporating a segmentation model was not considered for its limited precision in research stage. But with the optimization of the segmentation model, this AI model still has great potential in the future.



Figure 8.3-1 Segmentation Model <https://huggingface.co/spaces/hysts/ControlNet-v1-1>
the current segmentation model is not good at recognizing the details inside the main object, which can be seen from the window area in. The same is true for the texture design inside the swimsuit in the right image. [1]

Sport State (DINED)

During the early stages of user research, the topic of dynamic effects has consistently emerged as a significant aspect, with DINED users expressing a strong desire for such effects. However, upon conducting preliminary research, it became apparent that implementing dynamic processes poses significant challenges within the context of individual project research. It requires a collaborative team effort and a step-by-step development process. Consequently, the author considers this particular aspect as a recommendation for future research endeavors.

Are you willing to use the product I designed in your work ?

“Can't wait to”

Ohh that's what I what I like a lot here. I would like for such a tool is indeed. Maybe if I collect all the images online that I like right from different brands and then I could just put it in the model and then like just collect the ones that I really like and then generate.

8.3 Reference

[1] ControlNet V1.1 - a Hugging Face Space by hysts. (n.d.). <https://huggingface.co/spaces/hysts/ControlNet-v1-1>

Conclusion

In conclusion, the expert interview has yielded significant findings regarding the product under consideration.

1. First and foremost, both experts unanimously agree that the output tool DINED AI holds immense promise, offers an interesting AIGC method, and provides valuable assistance.
2. Furthermore, it is noteworthy that both experts express their willingness to use the product. However, when comparing the two experts, the individual engaged in design practice exhibits a particularly strong inclination towards adopting and utilizing the product.
3. The experts have also contributed a wealth of ideas and suggestions for the future development of the product.
4. From a design perspective, the expert engaged in design practice have provided detailed suggestions on enhancing user experience and optimizing functional design. His microscopic viewpoint of design operation serves as a valuable guide for future experience optimization. The results of this interview provide a reliable foundation for integrating the tool into all stages of the design process.
5. Moreover, the expert involved in design research have offered professional insights on integrating the product into various business models and expanding its application across industries. Her endorsement of design tool patterns not only reaffirms the research value of this project but also provides valuable macro-level advice for strategically positioning the tool- DINED AI.

Overall, the expert interview has provided a comprehensive understanding of the product's potential, the experts' willingness to use it, and a wide range of directions for future product development. The insights obtained from this interview serve as a robust basis for further advancements and strategic decisions related to the product.

Take Away

Key Evaluation

Likes



1. Meet current market demand
2. Have the potential to influence the iterative production and sales model



1. Super applicable for small design team
2. Able to assist with project plan presentation and sampling process

To Be Researched



1. Suitable for all ordinary users **VS** considering specially support current design habits of professional designers?

Dislikes



1. Not supportive for wetsuit design details generation
2. Can hardly convey the sport state of material's elasticity, thickness, and fit.....

Insight Keywords

Subdivision Target Group Type

1. Aimed for professional designers and inclusive for normal users
2. Small design team

Design Tool & Market business models

1. Fit in SHEIN-like KOL Marketing Model
2. Potential influence customized business model with DINED AI

Design Process

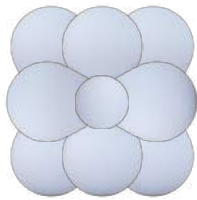
1. Demonstration before project approval
2. sampling (Prototyping)

UX Experience

1. Usage habits of professional designers
2. Small team collaboration
3. 3D animated function of DINED

09

Limitations, Recommendations, & Future Work



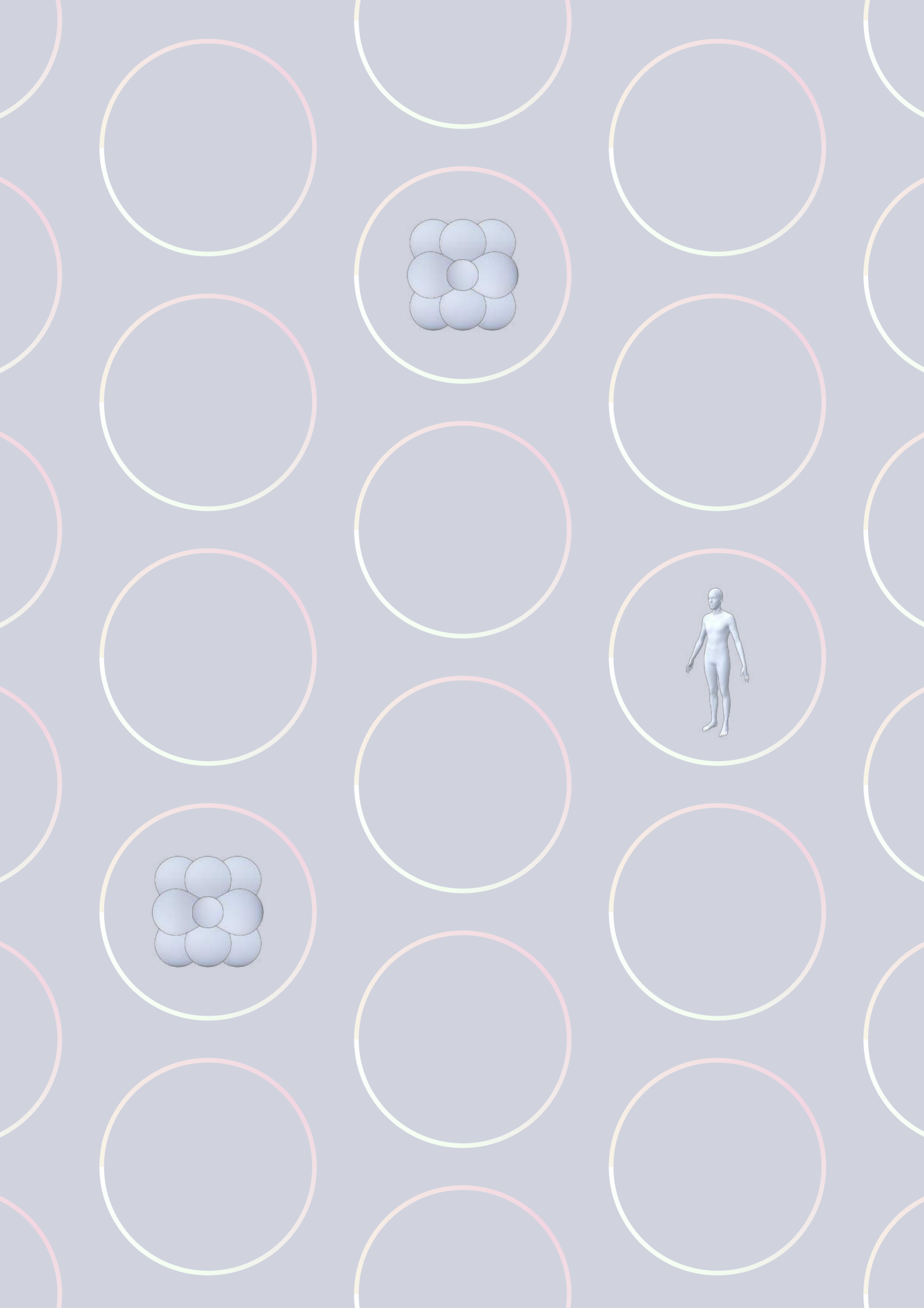
Chapter 9 serves as a critical examination of our research, shedding light on its inherent limitations, providing valuable recommendations, and charting a course for future exploration. By delving into the boundaries of this study, I gained a deeper understanding of its scope and context.

Chapter Overview

9.1 Limitations & Recommendations

9.2 Future Work

Take Away



9.1 Limitations & Recommendations:

Limitations of the method proposed in this project

Limitation 1-1: Situation where designing for diverse body shapes is not applicable: the loose version of the product design is not applicable

The method proposed in this project exhibits certain limitations, predominantly concerning its suitability for design processes closely intertwined with body shape. Although the approach demonstrates efficacy in crafting body-hugging designs, its effectiveness diminishes when applied to items that possess a distinct association with the body but do not conform tightly, such as loose-fitting garments.

The primary drawback of the proposed method lies in its limited applicability to designs that deviate from closely adhering to the contours of the human body. While it excels in generating intricate designs that accentuate the body shape and create visually stunning body-hugging ensembles, its utility is compromised when attempting to produce garments that prioritize a relaxed and unconstrained fit.

The intricacies arise when the method is employed to conceptualize loose-fitting garments, where the emphasis is on creating fluid and non-restrictive forms. The algorithm may struggle to accurately capture the desired aesthetics and nuances associated with these designs, as it heavily relies on the precise measurements and contours of the body. Consequently, the output may lack the desired volume, drape, and relaxed silhouette that are integral to loose-fitting apparel.



Figure 9.1-1 Image Dataset For Training the Loose-Fitting Garments LoRA



Figure 9.1-2 Negative Case: Image Outputs Of The Loose-Fitting Garments LoRA (Please See Appendix C9.1)

Moreover, the method's reliance on body shape as a foundational aspect introduces challenges in accommodating diverse body types and sizes. As the approach is inherently tied to the individual's unique proportions, it may encounter difficulties in adapting to varying body shapes, leading to suboptimal outcomes for individuals who do not conform to the assumed body model.

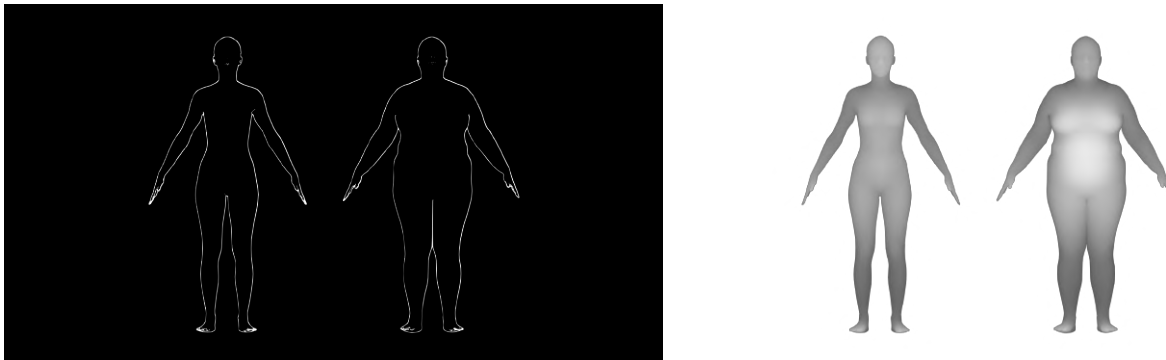


Figure 9.1-3 Use Open_Pose and Depth Map ControlNet Models

Furthermore, the proposed method's limitations extend to its practical implementation and feasibility. The computational complexity required to process intricate designs and generate accurate representations of loose-fitting garments might pose significant computational challenges. The method may require extensive computational resources and time-consuming calculations to achieve satisfactory results, potentially impeding its real-time application in design studios or manufacturing settings.

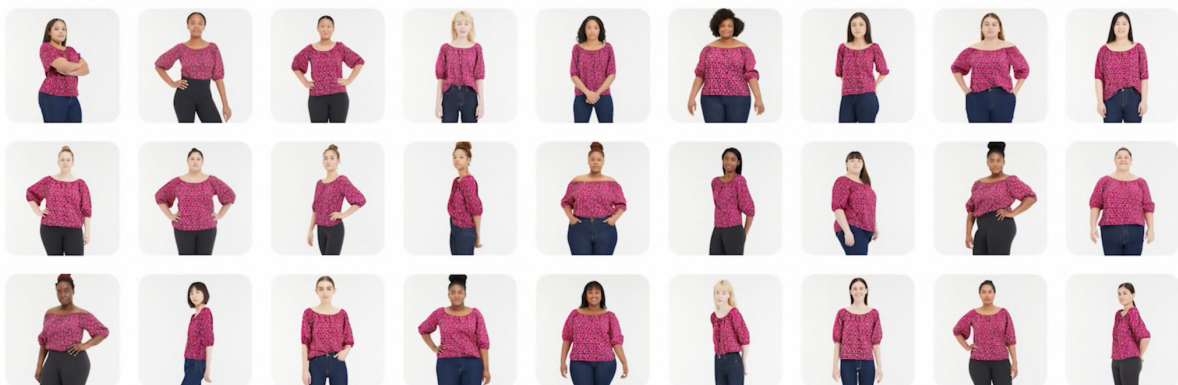


Figure 9.1-4 To Involve AI Algorithm Making Any Styles Working For Different Body Types In The Future e.g. Google's new AI-powered try-on feature, which taps generative AI to adapt clothing to different models.[1]

Recommendation 1-1: Further research and development are necessary.

Exploring alternative algorithms or hybrid approaches that incorporate additional design parameters beyond body shape could enhance the method's adaptability to various garment styles, including loose-fitting ones. Additionally, considering a broader range of body types and sizes during the training phase can improve the algorithm's ability to generate designs that cater to a more diverse population.

In summary, while the proposed method exhibits effectiveness in creating body-hugging designs, its limitations become apparent when applied to loose-fitting garments. The method's dependence on body shape and its potential computational challenges pose obstacles to its broader application. Addressing these limitations through continued research and innovation can pave the way for more versatile and inclusive design processes in the future.

9.1 Reference

[1] TechCrunch is part of the Yahoo family of brands. (2023, June 14). <https://techcrunch.com/2023/06/14/googles-new-generative-ai-lets-you-preview-clothes-on-different-models/>

Address the limitations or weaknesses in my methodology, data collection, or analysis that may have influenced the results or conclusions. And discuss the constraints or restrictions that affected the scope or generalizability of my findings.

Limitation 2-1: Only qualitative analysis was used for the evaluation

The current project is situated in the early stage of the first version within a regular software product development cycle. With its progress limited, one of the major challenges faced is the transformation of a mere new generative AI technology into a fully functional tool, essentially establishing from 0 to 1. During this critical stage, the primary objective of product design is to validate the significance of its existence, rather than prioritizing usability testing.

During this transformative phase, our team is dedicated to establishing a solid foundation for the product. As the author, I understand that the initial version may not be as user-friendly as we envision it to be in the future. Nonetheless, the main priority at this stage is to showcase the potential and demonstrate the unique capabilities of this generative AI technology.

By emphasizing the validation of its significance, the author aims to communicate the value proposition of the product to potential stakeholders, investors, and early adopters. The author wants to illustrate how this technology can revolutionize certain aspects of the industry or solve existing challenges in innovative ways. It is crucial to build a strong case for the product's potential impact and its ability to fill a gap in the market.

While usability testing is an integral part of product development, I acknowledge that focusing solely on usability at this stage would be premature. I believe that by first establishing the importance and impact of the product, we can garner support, gather feedback, and secure the necessary resources to further refine and enhance usability in subsequent iterations.

In summary, during this early stage of development, our project faces the challenge of transforming a nascent generative AI technology into a fully functional tool. The primary objective of product design is to validate the significance and value proposition of the product, rather than prioritizing usability testing. By establishing the importance of its existence, we aim to gain support and resources to further enhance usability in future iterations.

Recommendation 2-1: To interview more design experts and professional design practitioners , and Develop online prototype and collect data.

1. Expand the interview pool: In order to gather a wider range of insights and perspectives, it is recommended to conduct interviews with an increased number of design experts and professional practitioners. By doing so, a more comprehensive understanding of the subject matter can be achieved, resulting in more insightful and well-rounded recommendations.
2. Develop an online prototype for data collection: To bolster the credibility and applicability of the findings, it is recommended to create an interactive online prototype. This prototype will serve as a platform for data collection, enabling real-time feedback from users and stakeholders. By collecting data through the prototype, user behavior, preferences, and needs can be analyzed more effectively, contributing to more informed and tailored recommendations.

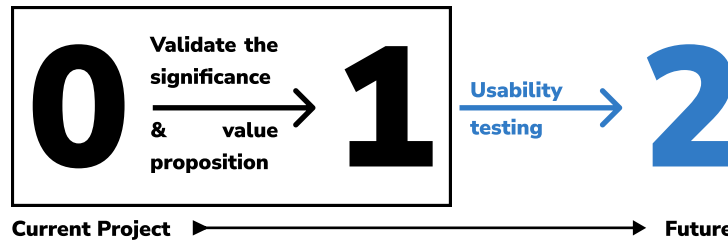


Figure 9.1-4 The current project is situated in the early stage of the first version

Limitation 2-2: The current availability of interview user data is limited in its scope.

The study encounters several challenges due to network limitations. One of the primary obstacles is the necessity to establish robust connections with industry practitioners in order to broaden our understanding and gain diverse perspectives. However, this task of finding a suitable interviewer can be quite challenging. On a positive note, there is a relatively larger pool of design students available for potential contact.

It is important to note, though, that only a smaller subset of these students possess prior experience working on projects that specifically focus on the human body. This limitation may impact the depth and breadth of insights we can gather from their involvement.

Additionally, another hurdle arises from the deployment of the prototype exclusively on local PCs. As a result, the process of inviting users for testing purposes becomes less straightforward than desired.

Recommendation 2-2: After inviting more design experts, the evaluation criteria of usability testing should be determined through interviews.

1. Foster collaboration within the design community: It is nice to establish a collaborative network with other researchers and design professionals. Engaging in discussions and knowledge-sharing activities will enable the researcher to tap into a wider pool of expertise, validate the recommendations with peers, and benefit from valuable insights and perspectives.
2. After inviting additional design experts, it is essential to establish the evaluation criteria for usability testing by conducting interviews. By expanding the scope of input from design experts, a broader range of perspectives and insights can be gained. These interviews will serve as a valuable source of information, enabling to refine and enhance the evaluation criteria for usability testing. This step is crucial to ensure that we capture a comprehensive understanding of user needs and preferences.
3. During the interviews, it can be posed that specific questions to the experts regarding their experiences, observations, and suggestions related to usability testing. By engaging in a meaningful dialogue, this project can gain valuable insights into the aspects that need to be considered while evaluating the usability of a design. These insights can range from user-friendliness and intuitive navigation to visual aesthetics and overall user satisfaction.

Acknowledge potential biases, sources of error, or external factors that could have impacted the outcomes.

Limitation 3-1: Given the rapid development of technology trends, it becomes challenging to predict whether the current popular methods will remain mainstream.

The field of stable diffusion and generative AI is experiencing rapid progress and continuous advancement. Within this project, the method in this project has been presented that stands out as an exceptionally cost-effective solution for now. This approach was created for its numerous advantages, including its open-source technology, easy accessibility, and the good ability to foster a thriving research community online. Although LoRA may not emerge as the most suitable approach in the immediate future, it is worth noting that influential companies like Google [1] have already made significant strides in introducing advanced AI models. This development serves as a compelling showcase of the immense potential for further advancements in the field. As the landscape of technology evolves, it is crucial to stay abreast of the latest innovations and trends, as they hold the key to unlocking even greater possibilities in the realm of SD technology and AI.

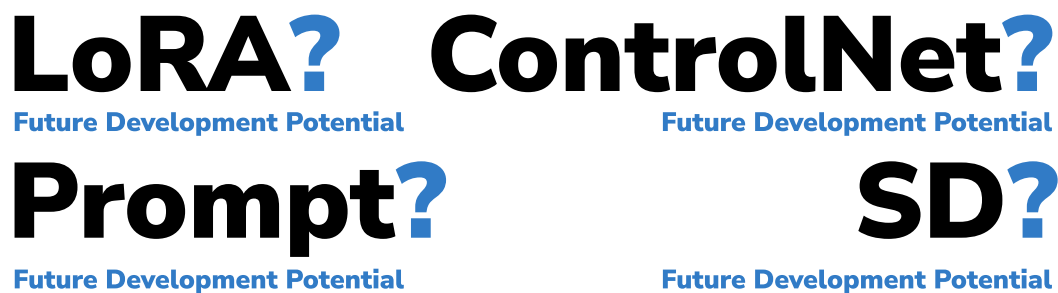


Figure 9.1-5 It is uncertain whether these technologies will continue to evolve and become useful in the future.

Recommendation 3-1: Keep following the development of generative AI technology

Stay updated with the ever-evolving landscape of generative AI technology, as its advancements continue to reshape the world as we know it. The field of generative AI, which encompasses various algorithms and models designed to generate creative and realistic outputs, holds immense potential for revolutionizing numerous industries and disciplines.

To stay abreast of the latest developments in generative AI, it is crucial to follow reputable research publications, and engage with the vibrant community of researchers and practitioners in the field. By doing so, you can gain a deeper understanding of the underlying techniques, learn from real-world applications, and potentially contribute to the advancements in generative AI yourself.

Limitation 3-2: There are few literatures on relevant applied research reference that can be cited.

The author incorporates references to technical discussions on popular platforms such as **Reddit** and **GitHub** to substantiate their arguments. This approach not only enriches the discussion but also lends credibility to their assertions. It is crucial, however, to recognize that these references are inherently limited to the specific context of each project being examined. Nonetheless, they offer valuable insights derived from the practical design experiences of individuals actively engaged in the field. Moreover, it is worth emphasizing that the references cited in this project have undergone validation and can be effectively applied within the unique context of the project itself. By drawing upon the collective wisdom and expertise of the online communities, the author harnesses a vast pool of knowledge to bolster their claims.

Additionally, it is essential to acknowledge that the field of AI technology is constantly evolving, and our understanding of its principles continues to deepen over time. As new developments unfold, there is a possibility of emerging principles that can bring about positive transformations. By remaining open to these advancements and incorporating them into our practices, we can ensure that our work aligns with the latest and most effective approaches.



Figure 9.1-6 LOGOs of Reddit and Github

Recommendation 3-2: Keep an eye on community discussions and the correctness of citations

It is essential to remain vigilant and actively monitor community discussions as well as ensure the accuracy and validity of the citations used. By keeping a close eye on ongoing conversations within relevant communities, such as online forums, social media platforms, and specialized groups, one can stay informed about the latest insights, debates, and advancements in a given field.

Community discussions serve as a valuable source of information, providing a platform for diverse perspectives and practical experiences. By actively participating in these conversations, individuals can gain a deeper understanding of the subject matter and identify potential nuances and nuances that may have been overlooked.

9.2 Future Work

Unexpected challenges & difficulties encountered during the project and their implications.

In the initial phase, the author dedicated significant time and effort to conducting extensive research on the technical capabilities of AI. Throughout this process, the author also came across various algorithms that offer improved results for fulfilling the project's objectives. Building upon this research, the author put forth optimization functions for enhancing DINED products, while acknowledging the need to address dynamic design issues. It is important to note that although numerous advanced algorithms were identified, their lack of open-source availability and absence of an online prototype posed certain challenges. Moreover, the author's personal proficiency in AI technology was limited. Consequently, despite the proposal of several DINED+AI methods during the early stage, they did not yield the desired success, and their effectiveness in practical applications could not be adequately verified.

Following the mid-term defense, the project took a different direction and embraced the ComfyUI method, which is currently widely recognized and favored for its simplicity in coding. This decision reflects the project's adaptability and responsiveness to prevailing trends in the field of AI, enabling the team to make notable progress in a positive direction.

Future Work For The Problem Above: Potential avenues for further exploration and investigation that emerged from this research but were beyond the scope of the current study.

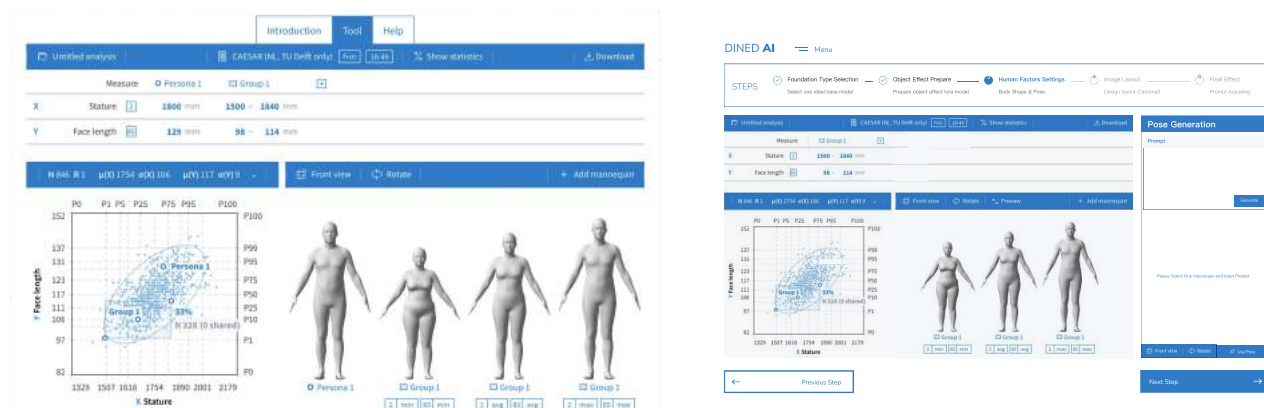
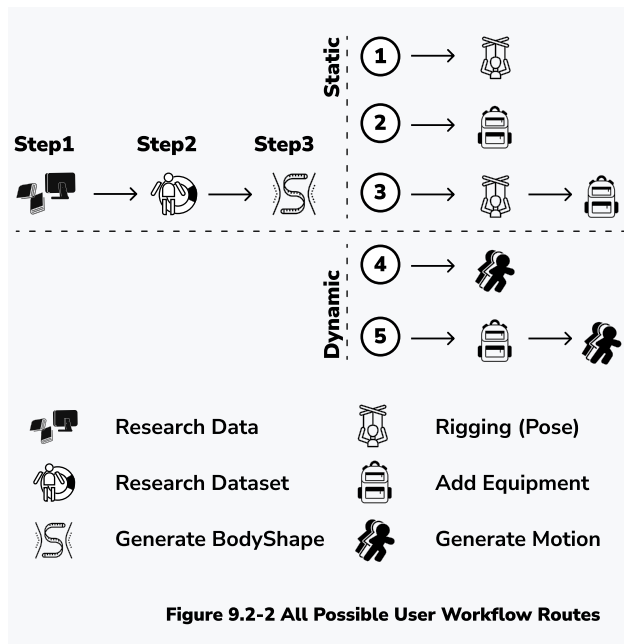


Figure 9.2-1 Seamless integration of DINED into DINED AI

Continuous optimization of DINED remains a key and unwavering focus for the future. The goals established in the pre-user study hold immense value and will continue to serve as our guiding principles as we forge ahead. One of the most notable advancements lies in the seamless integration of DINED into DINED AI, which holds the promise of substantial enhancements in the performance of the new product.

Through a meticulous optimization process, we are committed to ensuring that the final output not only meets but exceeds expectations, resulting in an exceptional user experience. This relentless pursuit of improvement involves a combination of rigorous testing, data analysis, and iterative refinement. By leveraging cutting-edge technologies and drawing upon the expertise of our talented team, we are confident in our ability to unlock the full potential of DINED and elevate it to new heights.

Future Optimisations in User Experience Map



Ideal Workflow

The User Experience Map outlines an ideal workflow in five steps:

Step 1 - Research data: Define the target group and determine the required anthropometric data.

Step 2 - Research dataset: Search for suitable databases or relevant literature.

Step 3 - Generate BodyShape: Create mannequins with different body types according to the requirements.

Step 4 - Rigging (Pose): Add poses to the mannequin.

Step 5 - Add equipment: Add equipment data to the mannequin.

Steps 1-3 are commonly shared steps, and users can choose their own user journey based on their expected outcome.

The expected outputs of the process for static data include acquiring poses, acquiring human body data with equipment, and equipment with poses.

The expected outputs of the process for dynamic data include acquiring human body motion, and equipment with human body motion.

Retain Users

Drawing on previous user research, the author has devised an optimal workflow for new users. The workflow commences with logging onto the platform, followed by gaining proficiency in using it, and culminating in utilizing the AIGC feature to generate anticipated outcomes. Based on user feedback on pain points, the workflow was scrutinized to identify the stages that could potentially dissuade users from continuing with the tool. Simultaneously, touchpoints were identified where the author hoped to retain users through the application's AIGC capabilities. The ultimate objective was to enable users to fulfill their needs primarily through the DINED tool, without having to switch between multiple software or platforms. This approach would help engender trust and reliance on the product among new users, building a positive reputation for the tool and making it the go-to solution for new users.

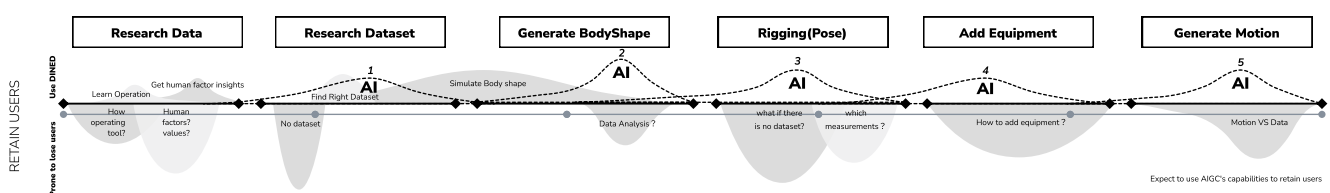
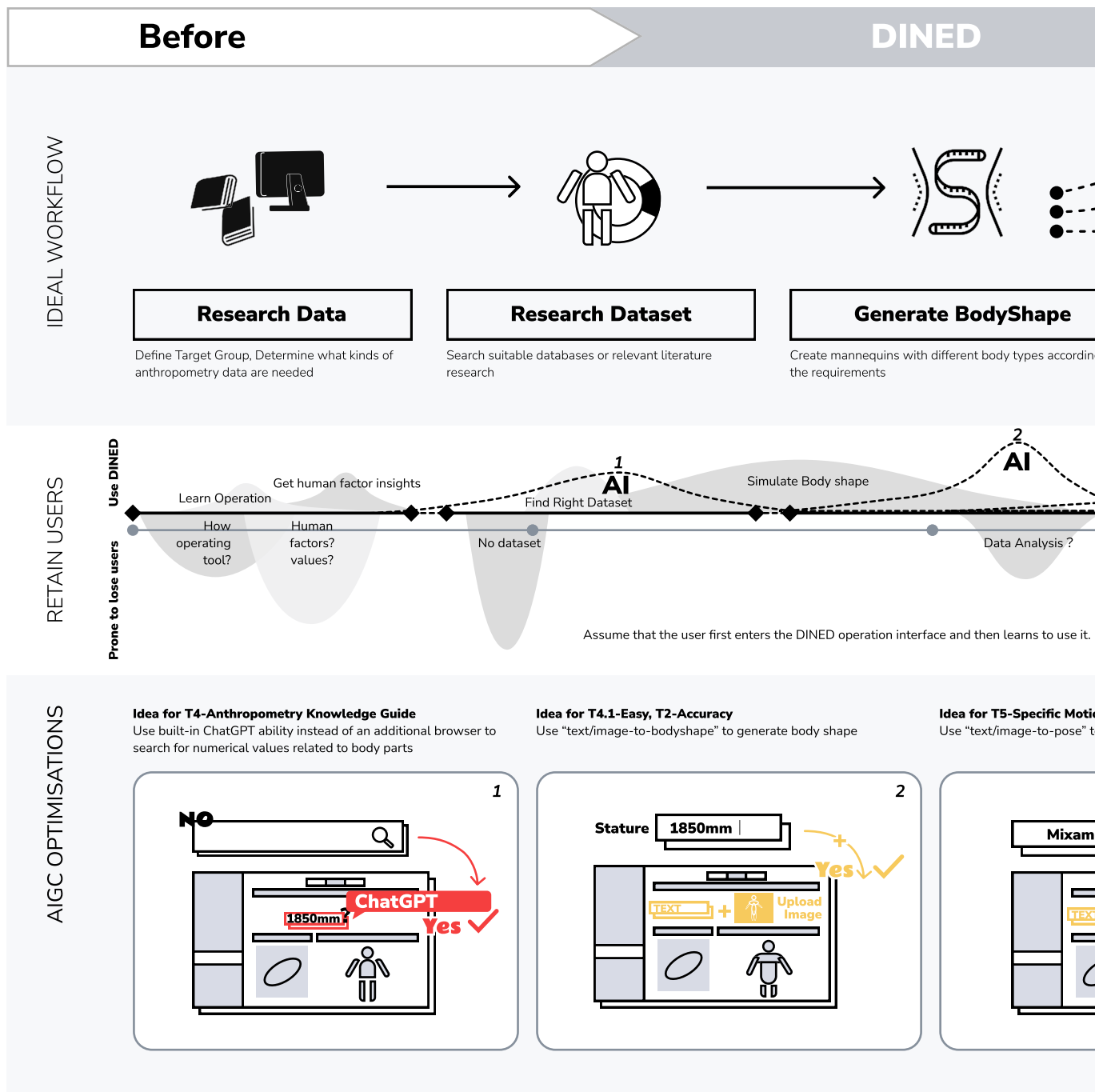


Figure 9.2-3 Retain Users

AIGC Optimisations in User Experience Map



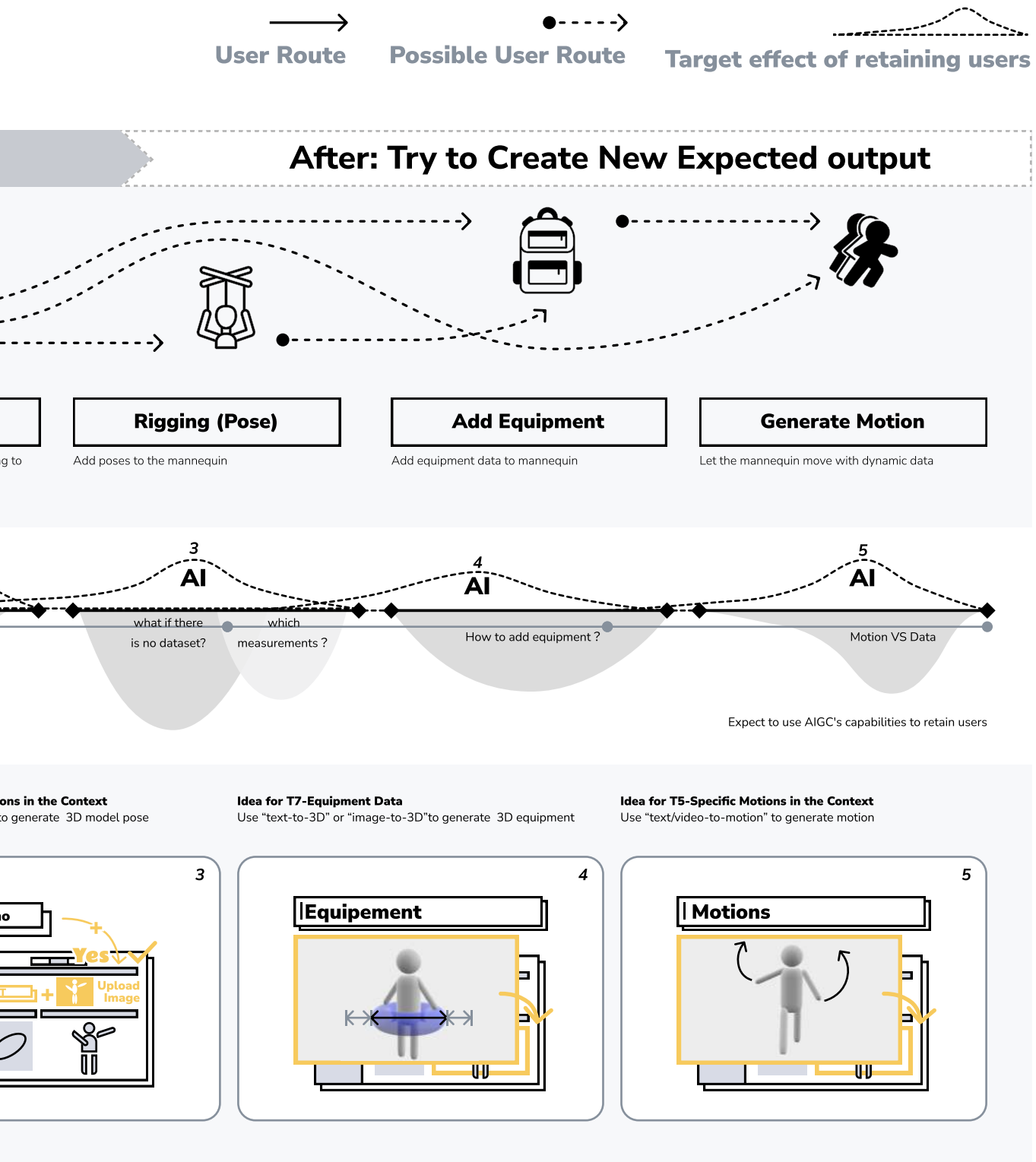


Figure 9.2-4 Experience Map—The figure illustrates the ideal workflow of DINED, highlighting the journey points where users are likely to drop off. To address this issue, five AIGC optimizations have been developed based on the analysis of user needs from the previous user interview. These optimizations aim to enhance the user experience and create a smoother journey with AIGC ability for users.

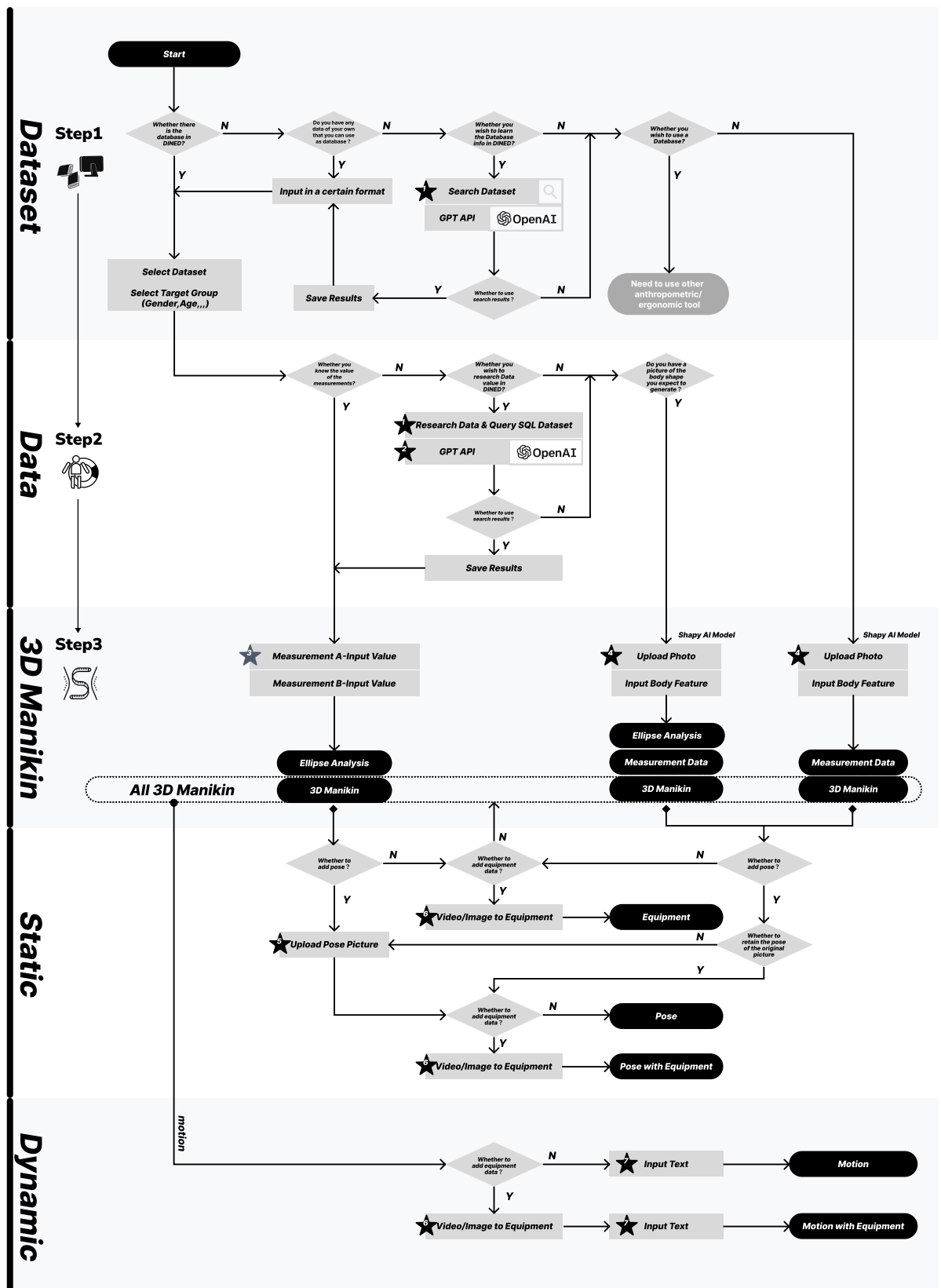
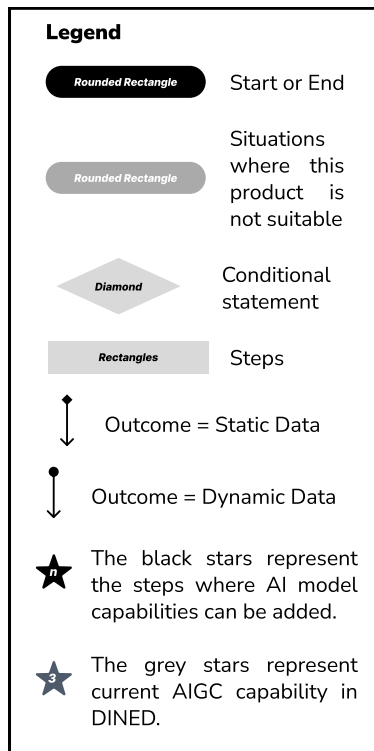
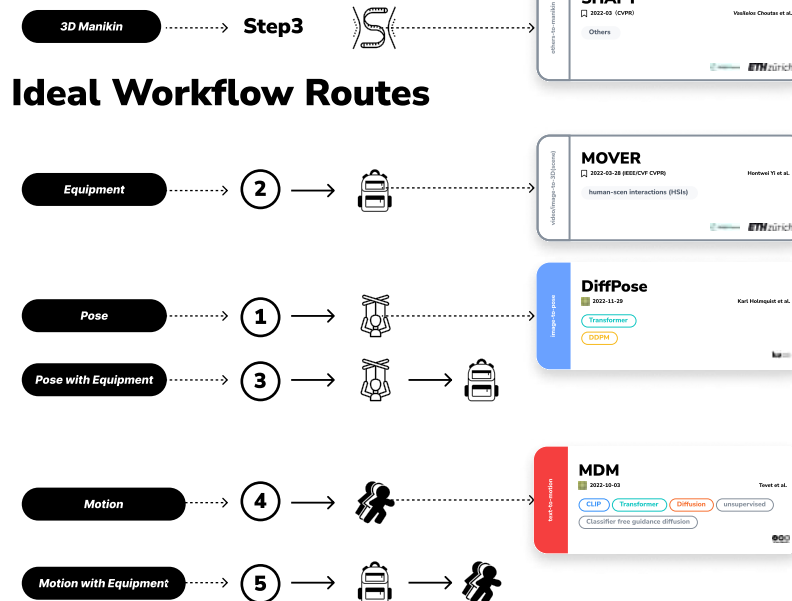


Figure 9.2-5 Flowchart—It is a diagram of the sequence of movements or actions of people or things involved in a complex system or activity. This diagram is the skeleton for my user flow.

AIGC Optimisations in Flowchart



Link Flowchart to Journey Map



Initial AI Vision for Added Capabilities

No.	Step	Brief How	Tool or AI Models	AIGC Pmtisations
★	Search Dataset GPT API OpenAI	Use GPT's powerful search and information integration functions to retrieve anthropometry data information	GPT	
★	Research Data & Query SQL Dataset GPT API OpenAI	Query using natural conversation via GPT API helps users find the data they need in DINED database quickly and intuitively	Bink Bink - Unlock Your Data's Potential (usebink.com)	
★	Upload Photo Input Body Feature	(Text+image)-to-manikin / SHAPY, a model that focuses on body shape and learns to predict dense 3D shape from a color image, well improving the accuracy of body shape generation	Shapy	
★	Upload Pose Picture	Image-to-pose / let the 3D manikin in a certain pose with image input	Diffpose (TBD)	
★	Video/Image to Equipment	Video/image-to-object/ let the pose or motion with equipment	Mover (TBD)	
★	Input Text	Text-to-motion / let the 3D manikin move with prompt. Motion Diffusion Model (MDM), a carefully adapted classifier-free diffusion-based generative model for the human motion domain.	MDM https://replicate.com/daanelson/motion_diffusion_model/examples	

Limitation

Weaknesses	Constraints
1-1 Situation where Designing for Diverse Body Shapes is Not Applicable: the loose version of the product design is not applicable	Limitations of the method proposed in this project primarily lie in its applicability to design processes closely tied to body shape. While the approach proves effective for creating body-hugging designs, it falls short when it comes to presenting items that have a certain connection to the body but are not form-fitting, such as loose-fitting garments.
2-1 Only qualitative analysis was used for the evaluation	Limited by the design phase (Early stage of first version) The challenge lies in creating a product from scratch, going from 0 to 1. At this stage, the primary focus of product design is to verify the value of its existence rather than conducting usability testing.
2-2 The current availability of interview user data is limited in its scope.	Network limitations present some challenges in this study. There is a need to establish stronger connections with industry practitioners to expand knowledge and perspectives. Finding a suitable interviewer can be a challenging task, although there is a relatively larger pool of design students that can be contacted. However, it should be noted that a smaller subset of these students have prior experience with projects specifically focused on the human body. Furthermore, due to the deployment of the prototype locally on PC , inviting users for testing purposes proves to be a less straightforward process.
3-1 Given the rapid development of technology trends , it becomes challenging to predict whether the current popular methods will remain mainstream due to their technical limitations.	The SD technology and AI technology are progressing rapidly. The method presented in this project stands out as an exceptionally cost-effective solution. It offers several advantages, such as being open source, easily accessible, and fostering a thriving research community. While LoRA may not be the most suitable approach in the near future, companies like Google are already making strides by introducing advanced AI models, showcasing the potential for further advancements.
3-2 There are few literatures on relevant applied research reference that can be cited.	The author includes references to technical discussions on Reddit and GitHub to support their points. While it's important to note that these results are limited to the specific context of each project, they still provide valuable insights derived from individuals' practical design experiences. It's worth mentioning that the references cited in this project have been validated and can be applied effectively within the project's own context. Furthermore, it's important to acknowledge that as our understanding of AI technology continues to deepen, there may be evolving principles that could bring about positive changes.

Suggestion

Suggest specific actions or interventions

More AI capabilities could be considered to add in DINED AI in the future

To interview more design experts and professional design practitioners /
Develop an online prototype and collect data

First, invite more experts to interview, and determine the evaluation criteria of usability testing through interviews.
Then quantitative analysis is carried out according to the standard, and then optimization is carried out

Continue to follow the development of generative AI technology

Keep an eye on community discussions and the correctness of citations

Take Away

Unexpected challenges & difficulties encountered during the project and their implications.

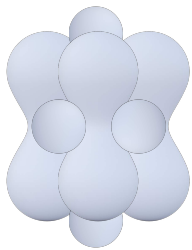
Initially, the author extensively researched AI capabilities and discovered various algorithms that could enhance DINED products. However, challenges arose due to limited access to advanced algorithms and the author's limited AI proficiency. As a result, proposed DINED+AI methods did not succeed practically. After the mid-term defense, the project shifted focus to the ComfyUI method, known for its simple coding and popularity. This adaptability led to notable progress in a positive direction.

Future Work

Continuous optimization of DINED remains a key focus for the future. The goals established in the pre-user study hold significant value and will continue to guide our efforts. Notably, as DINED is seamlessly integrated into DINED AI, we can anticipate substantial enhancements in the performance of the new product. This optimization process ensures that the final output meets and exceeds expectations, ultimately delivering an exceptional user experience.

10

Reflection



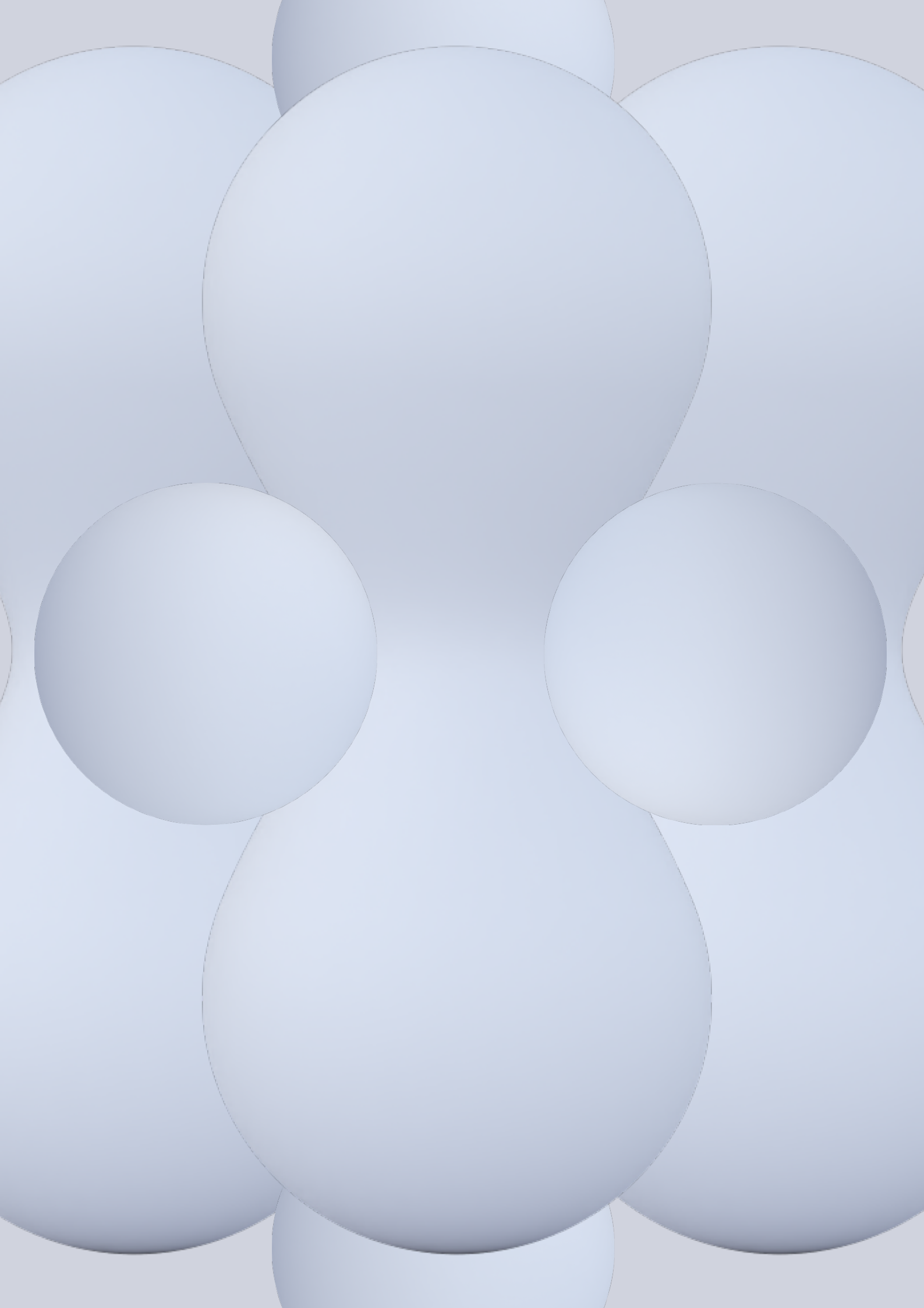
Chapter 10 discusses the design process and outcomes of this project to explore the AI potential future for DINED. This project aimed to explore the AI potential future for DINED that would encourage more DINED users or other designers to design for diverse body shapes easily .

Chapter Overview

10.1 Reflection

10.2 Supplement

Reference



10.1 Reflection

Introduction: In this reflection, I will discuss the design process and outcomes of my project to explore the AI potential future for DINED. This project aimed to explore the AI potential future for DINED that would encourage more DINED users or other designers to design for diverse body shapes easily .

Design Process: Throughout the design process, I followed a user-centered approach, beginning with explorative research to understand the needs and preferences of our DINED users. This research helped identifying the primary features and functionalities that AI solution for DINED needed to prioritize, such as knowledge guide, specific motions in the context, and equipment arrangement. This research phase greatly influenced the subsequent design decisions.

Successes: One of the main accomplishments of this design was the implementation of a visually concise and user-friendly interface. I focused on creating a simple design that allows users to easily navigate through different steps in generative AI. The utilization of LoRA model training and the DINED mannequin, which generates diverse real body shapes, improved the overall user experience in text-to-image AI learning. This means users do not need to switch between different graphical user interfaces (GUIs). The integration of the original DINED interface provided an easy win-win situation for the development of the new tool (DINED AI) and DINED. As users demand improvements to DINED, when DINED is refined, the new tool will be improved simultaneously. In addition, I have personally made significant progress in training the Lora model, drawing from the valuable insights gained from the community's practices.

Challenges: Throughout the project, I faced several significant challenges. One notable obstacle was the difficulty of identifying product opportunities amidst the overwhelming amount of AI technical information. Additionally, I encountered issues related to the divergence of AI directions around the mid-term, which required investing extra time and effort to ensure alignment between the goals of my mentors and myself. As a result, I continuously explored the intersection between potential users' perspectives and the technical aspects of the project. This experience greatly enhanced my design ability in developing products from a technological standpoint, which differs significantly from traditional human-centered interaction design. Currently, there is no definitive guidance available for this design process. Therefore, it becomes necessary to repeatedly fine-tune the product design positioning to adapt to the product's capabilities and the potential developments in AI technology.

Self-assessment of Degree of Completion

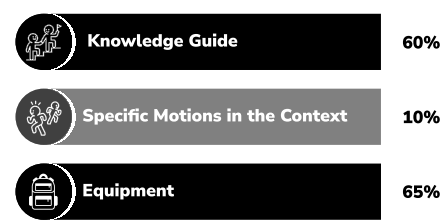


Figure 10.1-1 Self-assessment: the degree of completion of the final design to the ability to meet the top three needs of the users.

Despite the unsatisfactory completion of the self-assessment for the top 3 needs of DINED users, due to the adjustment of the implementation method, the author still strives to design according to the users' needs and lays the groundwork for future optimization.

Design Decision Analysis: The design decisions were driven by the goal of empowering DINED users to explore diverse body shapes using AI. I chose to integrate DINED into a new AI tool, enabling the text-to-image AI to generate content that accurately represents different body shapes based on anthropometry knowledge from DINED. Additionally, in the design of human-AI interaction, I consolidated all functions into a single tool and delegated complex parameter adjustments to backend presets. This approach aimed to reduce the learning curve of the text-to-image AI and enhance user engagement.

However, upon reflection, I realize that there could have been a better balance between minimizing the learning burden and allowing for professional fine-tuning. I acknowledge the need for a more delicate balance or involve soft prompt tuning [1] in future designs.

Lessons Learned: This design project has been instrumental in fostering my growth as an AI product designer. Throughout this journey, I have gained invaluable insights into the importance of conducting comprehensive technology and user research iteratively. Moreover, the obstacles I faced during the exploration phase have underscored the significance of early consideration of technical constraints in the design process. Looking ahead, I am determined to prioritize ongoing professional user research and foster collaborative partnerships with developers. By doing so, I aim to develop online prototypes that facilitate seamless implementation and enable efficient user testing in an online environment.

Conclusion: In conclusion, this design project has provided me with a valuable opportunity to apply my knowledge and skills in creating a groundbreaking AI product based on an existing professional tool. Throughout the process, I have encountered both successes and challenges, which have offered valuable insights into the iterative nature of AI product design.

I consider myself incredibly fortunate to have had the chance to work with this new AI technology during its infancy, and to have chosen it as the focal point of my graduation project, which represents the pinnacle of my postgraduate journey. I take great pride in the efficient method I have devised for diverse body shape design using generative AI, and I am thrilled to have received positive feedback from design experts. This project has closely mirrored the product development processes employed by companies, equipping me with essential skills for my future career as an AI product designer.

Looking ahead, I am committed to continuously refining my design skills and integrating the lessons I have learned from this project into my future endeavors. I believe that this experience has laid a solid foundation for my professional growth, enabling me to tackle new challenges with confidence and creativity.

10.1 Reference

[1] Wu, J., Yu, T., Wang, R., Song, Z., Zhang, R., Zhao, H., ... & Henao, R. (2023). InfoPrompt: Information-Theoretic Soft Prompt Tuning for Natural Language Understanding. arXiv preprint arXiv:2306.04933.

10.2 Supplement



Designer X AIGC Product

Insights on Developing AIGC Product

When considering incorporating AI capabilities into a product, the initial stages can be overwhelming. During this "fuzzy front period," it is crucial to have a clear understanding of several factors, such as why AI is necessary, the current state of AI capabilities, which AI capabilities can be applied, how to integrate AI algorithms into the product, and market trends for similar products. These and other concerns may leave development teams feeling uncertain.

In practical terms, product developers may initially launch functions based on their technical capabilities and then refine them based on user feedback and usage, as well as exploring revenue models. However, only larger companies typically have the financial and human resources to do so. For small and medium-sized enterprises, feasibility evaluations and design research are often conducted before taking action.

It is important to note that this project prioritizes user research before development action. By conducting thorough research and analysis, teams can ensure that their incorporation of AI capabilities aligns with user needs and expectations. This can ultimately lead to more successful product development and implementation.

Developing AIGC products as a designer has both advantages and challenges. The advantage is that, compared to programmers, designers pay more attention to the user experience of the product and are not restricted by a mindset focused solely on technical abilities. However, the challenge lies in understanding the capabilities of AI. Therefore, designers need to pay more attention to technological developments and read relevant literature. Articles in the AI field have been booming in recent years, covering everything from fundamental algorithms to top-level integrated application capabilities. The author of this article suggests that designers use Pattern Language Theory (PLT) as a way to quickly understand AI capabilities. With the support of hardware computing resources, designers only need to focus on the input and output of AI capabilities, as well as the advantages and differences between similar algorithms.

How to incorporate AI capabilities:

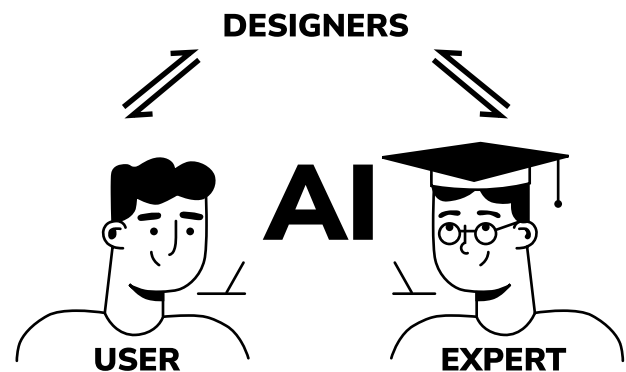
1. Designers should connect current and potential users with AI.

Understand user needs: This step is similar to the initial design research for regular software. Designers must focus on identifying the current usage difficulties of users and applying AI capabilities in the appropriate manner to solve problems, optimize processes, and enhance user experience.

Co-create with users: Co-creating AIGC products is more challenging compared to co-creating products and services. Accurately identifying the pain points of user context and finding appropriate solutions can be difficult, and designers must have excellent communication skills to convey AI capabilities to users. The PLT method can be useful in this regard, allowing designers to turn AI algorithms into probing toolkits such as cards for communication.

2. Designers must link experts with current and potential users.

Through literature research, designers can gain insights from experts in relevant fields, which can provide valuable inspiration for product design. By comparing "AI helping people" with "experts supporting people," designers can study the interaction between AI and humans to improve the design. One of the limitations of this project is the difficulty in contacting experts in relevant fields.



3. Designers need to start from two penetration points at the same time, which are user need and AI Technology

When embarking on the journey to explore AI opportunities for an existing product, the design process can be likened to excavating in two directions simultaneously. On one hand, delving deep into understanding user needs is essential, while on the other, digging into the realm of AI technology proves equally valuable. The perspective the designers gain from these two penetration points profoundly influences the design strategy of the new product. It's important to note that this process is iterative in nature, allowing us to refine and enhance our approach continually.

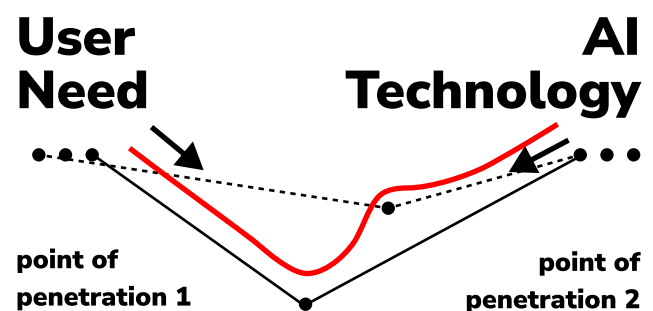


Figure 10.2-2 Iterate from two entry points

Chapter 10

Reference (4)

Chapter 8

8.2 Reference

[1] Decoder. (n.d.). How did SHEIN become the top fashion seller of China's cross-border e-commerce? [www.linkedin.com](https://www.linkedin.com/pulse/how-did-shein-become-top-fashion-seller-chinas-cross-border-). <https://www.linkedin.com/pulse/how-did-shein-become-top-fashion-seller-chinas-cross-border->

8.3 Reference

[1] ControlNet V1.1 - a Hugging Face Space by hysts. (n.d.). <https://huggingface.co/spaces/hysts/ControlNet-v1-1>

Chapter 9

9.1 Reference

[1] TechCrunch is part of the Yahoo family of brands. (2023, June 14). <https://techcrunch.com/2023/06/14/googles-new-generative-ai-lets-you-preview-clothes-on-different-models/>

Chapter 10

10.1 Reference

[1] Wu, J., Yu, T., Wang, R., Song, Z., Zhang, R., Zhao, H., ... & Henao, R. (2023). InfoPrompt: Information-Theoretic Soft Prompt Tuning for Natural Language Understanding. arXiv preprint arXiv:2306.04933.

Thanks !

