# Chat-to-Design: AI Assisted Personalized Fashion Design

Weiming Zhuang[1,4] Chongjie Ye[2] Ying Xu Pengzhi Mao Yonggang Wen[3]
[1]S-Lab, Nanyang Technological University [2]The Chinese University of Hong Kong
[3]Nanyang Technological University [4]SenseTime Research
weiming001@e.ntu.edu.sg,ygwen@ntu.edu.sg

## ABSTRACT

In this demo, we present Chat-to-Design, a new multimodal interaction system for personlized fashion design. Compared to traditional system that recommands apparels based on keywords, Chat-to-Design enable users to design clothes in two steps: 1) coarse-grained selection via conversation and 2) fine-grained adjustment via interative interface. It encompasses three sub-systems to deliver immersive user experience: A conversation system empowered by natual language understanding to accept users requests and manages the dialog; A multimodal fashion retrieval system empowered by a large-scale language-image pretrained network to retrieve users' desired apparels; A fashion design system empowered by emerging generative techniques to edit attributes of selected clothes.

## CCS CONCEPTS

• **Computing methodologies** → **Computer vision**; **Natural language processing**; • **Human-centered computing** → **Human computer interaction (HCI)**; • **Information systems** → *Information retrieval.*

## KEYWORDS

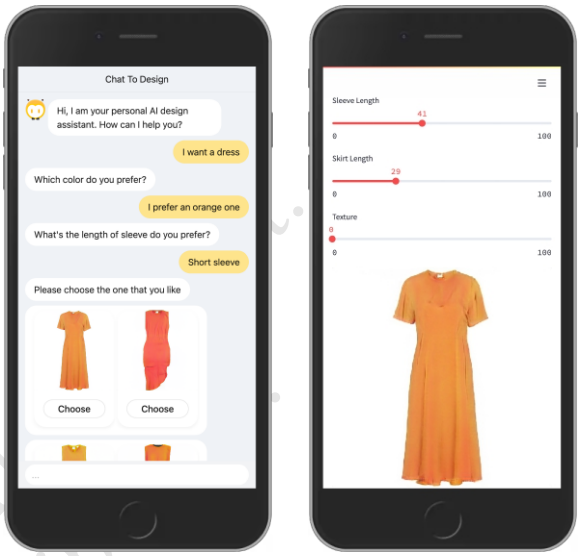
Apparel design, apparel editing, multimodal fashion retrieval

**ACM Reference Format:**
Weiming Zhuang[1,4] Chongjie Ye[2] Ying Xu Pengzhi Mao Yonggang Wen[3]. 2022. Chat-to-Design: AI Assisted Personalized Fashion Design. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

Searching for desired clothes to buy on e-commerce websites has evolved into a new habit of the new generation. However, classic recommendation systems are difficult to fully capture users' preferences based on clicking or rating data [2]. This demo presents a new multimodal interaction system for personalized fashion design, Chat-to-Design. It understands users' requests via conversation and enables them to further customize clothes.

The user-friendly interface of Chat-to-Design encompasses a complex system that integrates *conversation system* empowered by natural language understanding and dialog system, *multimodal fashion retrieval system* empowered by a large-scale language-image pre-trained network, and *fashion design system* empowered by the domain-specific generative networks. Chat-to-Design helps design users favored cloth in two-step: coarse-grained selection via conversation in Figure 1a and fine-grained adjustment via interactive interface in Figure 1b. To the best of our knowledge, we are the first to explore such a comprehensive multimodal interaction system for design in the multimedia area.




(a) Chat (b) Design

**Figure 1: User interfaces and illustrations of Chat-to-Design.**

Our system is different from the existing system in many different aspects. Different from traditional keyword-based or filter-based search systems, Chat-to-Design employs dialog to interpret the needs of users. Different from a pure chatbot [13] or pure multimodal fashion retrieval system [8, 14], Chat-to-Design not only provides a dialog system but also embraces multimodal interactions. Different from existing method on AI-assisted fashion design [4, 12], Chat-to-Design builds an end-to-end multimodal interaction system and leverages the lastest generative approach, StyleGAN [6] StyleFlow [1] to enable modifying apparel based on attributes.

## 2 SYSTEM ARCHITECTURE

The system is designed to be user-friendly and real-time interactive. Figure 2 depicts the system architecture of Chat-to-Design. It comprises three sub-systems: *conversation system* accepts and interprets users' requests, manages the dialog, and sends requests to the other two systems; *multimodal fashion retrieval system* retrieves the most


*Conference'17, July 2017, Washington, DC, USA*

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

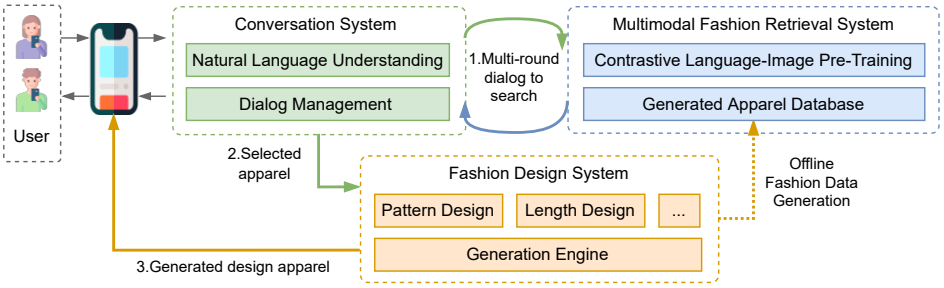



**Figure 2: The system architecture of Chat-to-Design. It includes three sub-systems: conversation system, multimodal fashion retrieval system, and fashion design system.**

appropriate apparel according to description; *fashion design system* enables users to further edit the details of the selected apparel. Next, we introduce these three systems in detail.

## 2.1 Conversation System

The conversation system is the entrance of the whole system. It consists of a natural language understanding (NLU) module and a dialog management module. NLU has been a well-studied research field. We adopt the widely-used MitieNLP [7] toolkit for intent recognition and entity extraction inside the NLU module. To equip the NLU module with domain knowledge of fashion, we generate synthesized fashion dialog data using Chatito [9]. The dialog management manages multi-round conversations. We integrate these two modules using the widely-adopted Rasa [11] framework. The conversation system responds with multimodal contents: text, images, and actions, which greatly enriches users' experience.

## 2.2 Multimodal Fashion Retrieval System

Multimodal fashion retrieval system searches and returns suitable clothes with given extracted user requests. Multimodal search is attracting considerable attention. We build this system based on CLIP (Contrastive Language-Image Pre-Training) [10], which is pretrained on 400 million pairs of images and text pairs from the Internet. CLIP is powerful in establishing the connection between image and text. We further empower it with domain knowledge of fashion by fine-tuning its pretrained Vision Transformer (ViT-B/32) [3] using Fashion200K dataset [5], which contains 200K fashion image and text pairs.

## 2.3 Fashion Design System

We support real-time fashion attribute editing in the fashion design system. It allows users to flexibly design clothes by adjusting attributes like texture and sleeve length. The system is empowered by the latest generative techniques, including StyleGAN [6] and StyleFlow [1]. To the best of our knowledge, we are the first to adopt both StyleGAN and StyleFlow to the fashion domain. We train these models with a newly constructed large-scale fashion dataset to enable them with knowledge in the fashion domain. In particular, we construct the dataset by first collecting around 100,000 images of 6 attributes and merging it with Fashion200K [5] data. Then, we obtain clean the datasets by removing the background and human bodies with semantic segmentation and edge detection techniques.

**(a) Clothes generated using existing datasets**

**(b) Clothes generated using our newly constructed dataset**

**Figure 3: Generated cloth samples using different datasets.**

Figure 3 illustrates that clothes generated using our clean dataset are superior to using existing datasets like Fashion200K. Besides, we generate fashion data offline and merge it into the database for multimodal fashion retrieval.

# 3 DEMONSTRATION

Figure 1 demonstrates the user interfaces and workflow of interacting with the system. Chat-to-Design helps design users favored cloth in two steps: coarse-grained selection via conversation in Figure 1a and fine-grained adjustment via interactive interface in Figure 1b. Firstly, the user starts a conversation with a query sentence like "I want a dress". The system chats with the user to collect other desired attributes of the dress, such as color and sleeve length. After that, the system suggests multiple suitable dresses, which the user can either select the desired dress or provide further description. After selecting a dress, the user can further tune the details of attributes of clothes, such as adjusting the textures and changing sleeve length. Chat-to-Design enables the user to view the effect of these changes immediately and confirm the desired design.

Chat-to-Design has a broad scope of potential applications. It can be integrated into e-commerce websites to better capture consumers' needs and support the customer-driven manufacturing process. Besides, by further integrating virtual try-it-on, users can share their photos wearing self-designed clothes on social media and short videos. Lastly, Chat-to-Design has the potential to support user-designed virtual apparel in the metaverse.

# REFERENCES


[1] Rameen Abdal, Peihao Zhu, Niloy J Mitra, and Peter Wonka. 2021. Styleflow: Attribute-conditioned exploration of stylegan-generated images using conditional continuous normalizing flows. *ACM Transactions on Graphics (TOG)* 40, 3 (2021), 1–21.

[2] Robert M Bell and Yehuda Koren. 2007. Lessons from the netflix prize challenge. *Acm Sigkdd Explorations Newsletter* 9, 2 (2007), 75–79.

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*. https://openreview.net/forum?id=YicbFdNTTy

[4] Alpana Dubey, Nitish Bhardwaj, Kumar Abhinav, Suma Mani Kuriakose, Sakshi Jain, and Veenu Arora. 2020. AI Assisted Apparel Design. *arXiv preprint arXiv:2007.04950* (2020).

[5] Xintong Han, Zuxuan Wu, Phoenix X Huang, Xiao Zhang, Menglong Zhu, Yuan Li, Yang Zhao, and Larry S Davis. 2017. Automatic spatially-aware fashion concept discovery. In *Proceedings of the IEEE international conference on computer vision*. 1463–1471.

[6] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4401–4410.

[7] Davis E. King. 2009. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research* 10 (2009), 1755–1758.

[8] Lizi Liao, Xiangnan He, Bo Zhao, Chong-Wah Ngo, and Tat-Seng Chua. 2018. Interpretable multimodal retrieval for fashion products. In *Proceedings of the 26th ACM international conference on Multimedia*. 1571–1579.

[9] Rodrigo Pimentel. 2019. *Chatito*. https://github.com/rodrigopivi/Chatito

[10] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*. PMLR, 8748–8763.

[11] Rasa. 2021. *Rasa Open Source*. https://github.com/RasaHQ/rasa

[12] Negar Rostamzadeh, Seyedarian Hosseini, Thomas Boquet, Wojciech Stokowiec, Ying Zhang, Christian Jauvin, and Chris Pal. 2018. Fashion-gen: The generative fashion dataset and challenge. *arXiv preprint arXiv:1806.08317* (2018).

[13] Hui Wu, Yupeng Gao, Xiaoxiao Guo, Ziad Al-Halah, Steven Rennie, Kristen Grauman, and Rogerio Feris. 2021. Fashion iq: A new dataset towards retrieving images by natural language feedback. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11307–11317.

[14] Hui Wu, Yupeng Gao, Xiaoxiao Guo, Ziad Al-Halah, Steven Rennie, Kristen Grauman, and Rogerio Feris. 2021. Fashion iq: A new dataset towards retrieving images by natural language feedback. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11307–11317.