ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR904

Conversational Fashion Outfit Generator Powered by GenAI

Deepak Gupta[1]
Computer Department
Dr. D Y Patil Institute of Engineering
Management and Research
Pune Maharashtra, India

Harsh Ranjan Jha[2]
Computer Department
Dr. D Y Patil Institute of Engineering
Management and Research
Pune Maharashtra, India

Maithili Chhallani[3]
Computer Department
Dr. D Y Patil Institute of Engineering
Management and Research
Pune Maharashtra, India

Mahima Thakar[4]
Computer Department
Dr. D Y Patil Institute of Engineering
Management and Research
Pune, Maharashtra, India

Dr. Amol Dhakne[5] (Associate Professor)
Computer Department
Dr. D Y Patil Institute of Engineering
Management and Research
Pune Maharashtra, India

Prathamesh Parit[6]
AI/DS Department
Dr. D Y Patil Institute of Engineering
Management and Research
Pune, Maharashtra, India

Hrushikesh Kachgunde[7]
AI/DS Department
Dr.D Y Patil Institute of Engineering
Management and Research
Pune, Maharashtra, India

Abstract:- The convergence of artificial intelligence and fashion has given rise to innovative solutions that cater to the ever-evolving needs and preferences of fashion enthusiasts. This report delves into the methodology behind the development of a "Conversational Fashion Outfit Generator powered by GenAI," an advanced application that leverages the capabilities of Generative Artificial Intelligence (GenAI) to create personalized fashion outfits through natural language interactions. The model outlines the essential elements of the methodology, including data collection, natural language understanding, computer vision integration, and deep learning algorithms. Data collection forms the bedrock, as access to a diverse dataset of fashion-related information is critical for training and fine-tuning AI models. Natural Language Understanding (NLU) is instrumental in comprehending user input and generating context-aware responses, ensuring meaningful and engaging conversations. Computer vision technology is integrated to analyze fashion images, recognizing clothing items, styles, and colors, thus aiding in outfit recommendations. Deep learning algorithms, particularly recurrent and transformer-based models, form the backbone of the system, generating personalized and contextually relevant fashion suggestions. This methodology not only underpins the "Conversational Fashion Outfit Generator" but also reflects the evolving landscape of AI in the fashion industry, where personalized, interactive experiences are becoming increasingly paramount in the realm of fashion and e-commerce.

Keywords:- *Generative AI, Stable Diffusion, Warp Model, Fashion Recommendation.*

I. INTRODUCTION

A "Conversational Fashion Outfit Generator powered by GenAI" is an innovative application that leverages the capabilities of Generative Artificial Intelligence (GenAI) to assist users in creating personalized fashion outfits through natural language conversations. This technology represents a convergence of AI-driven fashion recommendation and conversational AI, offering a dynamic and interactive solution for fashion enthusiasts. By engaging in a dialogue with the system, users can describe their preferences, occasions, and styles, and in response, the AI system generates tailored fashion outfit suggestions. This approach aims to enhance the user's shopping and styling experience by providing real-time, AI-generated outfit recommendations that align with individual tastes and needs.

The "Conversational Fashion Outfit Generator powered by GenAI" reflects the evolving landscape of applications in the AI fashion industry, catering to both fashion-conscious consumers and e-commerce platforms seeking to provide highly personalized and engaging shopping experiences. We explore the intricate methodology that underlies the development of the "Conversational Fashion Outfit Generator powered by GenAI." Each component of the methodology - data collection, natural language understanding, computer vision integration, and deep learning algorithms - plays a pivotal role in creating

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR904

this sophisticated AI-driven fashion platform. Together, they allow for the interpretation of user preferences, the analysis of fashion images, and the generation of personalized outfit recommendations.

Beyond its technical intricacies, this technology represents a significant advancement in the fashion and e-commerce landscape, providing highly personalized and engaging shopping experiences that resonate with today's fashion-conscious consumers. This report aims to elucidate the methodology that powers this innovative solution and shed light on the emerging paradigm of conversational AI in the realm of fashion.

The outfit generator is powered by advanced machine learning algorithms that analyze various fashion trends, colors, and patterns to create unique and eye-catching ensembles. Simply input your desired outfit type, color scheme, and any specific items you'd like to include, and our AI system will generate a variety of options for you to choose from. 0t f2.

With the conversational AI outfit generator, you can save time and energy when it comes to choosing the perfect outfit. Say goodbye to the endless scrolling through countless fashion websites and hello to a personalized, efficient, and enjoyable shopping experience.

II. EASE OF USE

This is the software configuration in which the project was shaped. The programming language used, tools used, etc are described here.

- Python
- Pytorch
- JavaScript
- HTML , CSS
- For data- NumPy and pandas package

➢ *Back-end*

This is working of backend development which is configuration , administration and management of databases and servers.

- MongoDB- Used as a backend data
- Gradio
- JavaScript
- Flask

III. METHODOLOGIES

A. Stable Diffusion

➢ *Introduction*

Stable Diffusion is an AI model that creates unique visual effects based on text and image cues. It was released in 2022. You can use this model to create videos and animations as well as images. The model is based on the propagation process and uses the latent space.

This reduces performance and you can model on a desktop or laptop equipped with a GPU. Stable expansion thanks to flexible learning can be customized to meet your specific needs with just five frames. As a diffusion model, steady diffusion differs from most other diffusion models. In principle, the diffusion model uses Gaussian noise to encode the image.

They then used sound prediction and back-and-forth techniques to create images. Besides the difference of having a diffusion pattern, stable diffusion is also unique in that it does not use the pixel space of the image. Instead it uses a simplified hidden field

➢ *Training*

Given the input image z0, the image propagation algorithm gradually adds noise to the image and produces an image noise zt; where t represents the noise addition time. Given the conditions including time step t, index ct, and specific function cf, the image propagation algorithm learns the Ïµî¸ network to estimate the noise added to the noise image zt; where L

Hi

$$= \mathrm{E}\boldsymbol{z}\ ,\boldsymbol{t},\boldsymbol{c}\ ,\boldsymbol{c}\ ,\epsilon{\sim}\mathrm{N}(0,1) \quad \|\epsilon - \epsilon\theta(\boldsymbol{z}t,\ \boldsymbol{t},\ \boldsymbol{c}t,\ \boldsymbol{c}\mathrm{f}))\|^2$$

Where is the overall learning objective of where is the general learning target of all difference-fusion models. This learning objective is applied directly to accurate diffusion models using ControlNet. This approach improves ControlNet's ability to directly identify details (e.g. edges, pose, depth, etc.) in image data that are converted into cues. but the model should always be able to predict the shape well.

Instead of quickly learning the management, we observe that the model is suddenly completed according to the output image many times; We call this the “sudden convergence phenomenon” as shown in Figure 4.”.

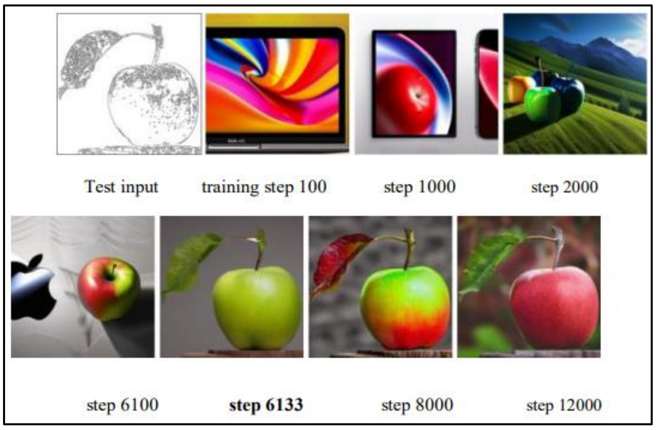

Fig 1 The phenomenon of sudden convergence gives an idea. Due to zero convolution, ControlNet consistently predicts image quality throughout the training process. At some steps of the training process (for example, step 6133 marked in bold), the model suddenly knows how to follow the instructions.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR904

Fig 2 Classifier-Free Guidance (CFG) and the proposed CFG Resolution Weighting (CFG-RW)

➢ *Conclusion*

The phenomenon of stable diffusion in AI marks a significant milestone in the integration of advanced technologies into our daily lives and industries. As AI continues to mature and evolve, we observe steady and pervasive dissemination of AI applications across diverse sectors such as healthcare, finance, transportation, manufacturing, and beyond. This stable diffusion is driven by the growing recognition of AI's transformative potential to enhance efficiency, productivity, and decision-making processes. However, challenges such as ethical considerations, bias mitigation, and regulatory frameworks must be addressed to ensure responsible and equitable deployment of AI technologies. Overall, the phenomenon of stable diffusion underscores the enduring impact of AI on reshaping societies, economies, and the future of work, heralding a new era of innovation and opportunity."

B. Warp Model

➢ *Introduction*

Virtual try-on technology has garnered significant attention in research due to its potential to revolutionize consumers' shopping experiences. This innovative technique aims to seamlessly transfer clothing from one image to another, creating a realistic composite image that enhances the visualization of how garments would look on a specific individual. Central to this task is the need to ensure that synthetic results are convincingly realistic while preserving the textural details of the clothing and other characteristic attributes of the target person, such as appearance and pose.

Many prior studies have leveraged Generative Adversarial Networks (GANs) as the foundation for generating lifelike images. Additionally, to further enhance the fidelity of results, some researchers have integrated explicit warping modules. These modules align the target clothing with the contours of the human body before feeding it into the generator along with a clothes-agnostic image of the person, yielding the final try-on result. Some advancements have also extended this task to high-resolution scenarios. However, the efficacy of such frameworks is heavily reliant on the quality of the warped garments. Subpar warping can hinder the generation of faithful results. Moreover, GANs-based generators inherit certain weaknesses of the GAN model itself, such as convergence issues related to hyperparameter selection and mode drop in the output distribution. Despite the promising outcomes of these approaches, it's crucial to address these limitations to further improve the reliability and performance of virtual try-on systems.

➢ *Diffusion Model*

A novel approach in image generation, Denoising Diffusion Probabilistic Models (DDPM) [19, 39], has emerged, offering the capability to generate realistic images from a normal distribution by reversing a gradual noising process. Despite its potential to generate diverse and realistic images, DDPM suffers from slow sampling speeds, limiting its widespread application. Addressing this limitation, Song et al. [40] introduced DDIM, which transforms the sampling process into a non-Markovian process, enabling faster and deterministic sampling. In parallel efforts to reduce computational complexity and resource requirements, latent diffusion models (LDM) [35] have been developed. LDM utilizes a set of frozen encoder-decoder pairs to conduct the diffusion and denoising processes in the latent space. As the diffusion model continues to evolve and mature, it has become a formidable contender to Generative Adversarial Networks (GANs) in image generation tasks.

Simultaneously, researchers are exploring methods to enhance control over diffusion model generation. Integrating text-to-image technology, several studies [34, 35, 37] have incorporated text information as a conditioning factor in the denoising process. This guides the model to generate images that align with the provided text descriptions. Techniques such as ILVR [27] and SDEdit [5] intervene in the denoising process at the spatial level, providing additional control over diffusion model generation.

Furthermore, recent advancements [31, 46] have focused on facilitating the transfer of diffusion models to different tasks, enhancing their versatility and applicability. Despite these developments, challenges remain in achieving optimal performance and versatility in diffusion model-based image generation.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR904

IV. DIAGRAMS

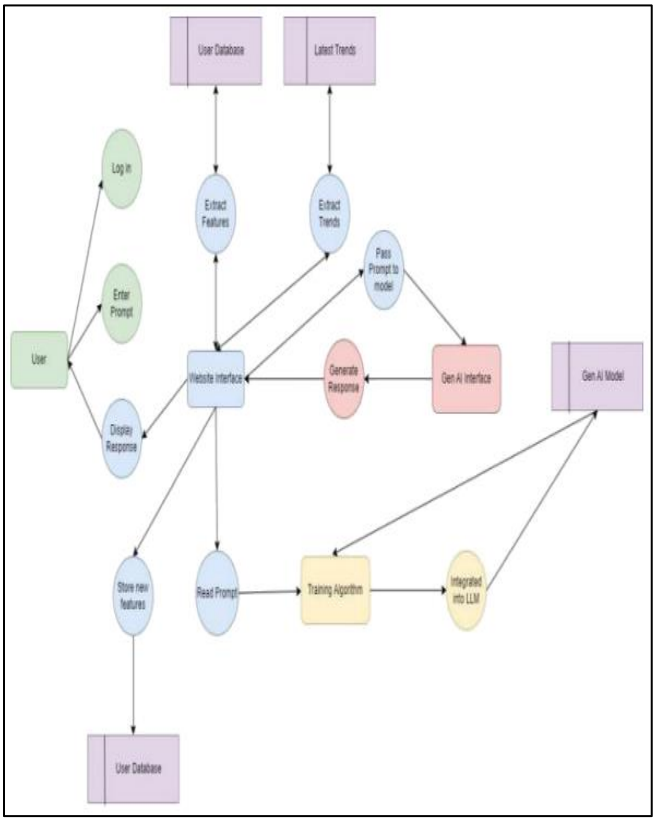

Fig 3 Data Flow Diagram

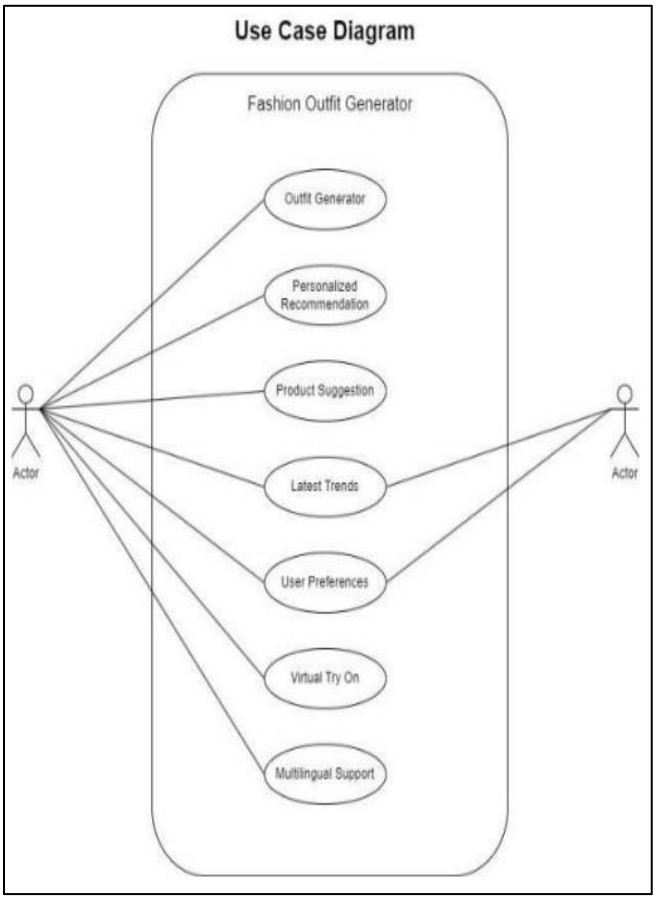

Fig 4 Use Case Diagram

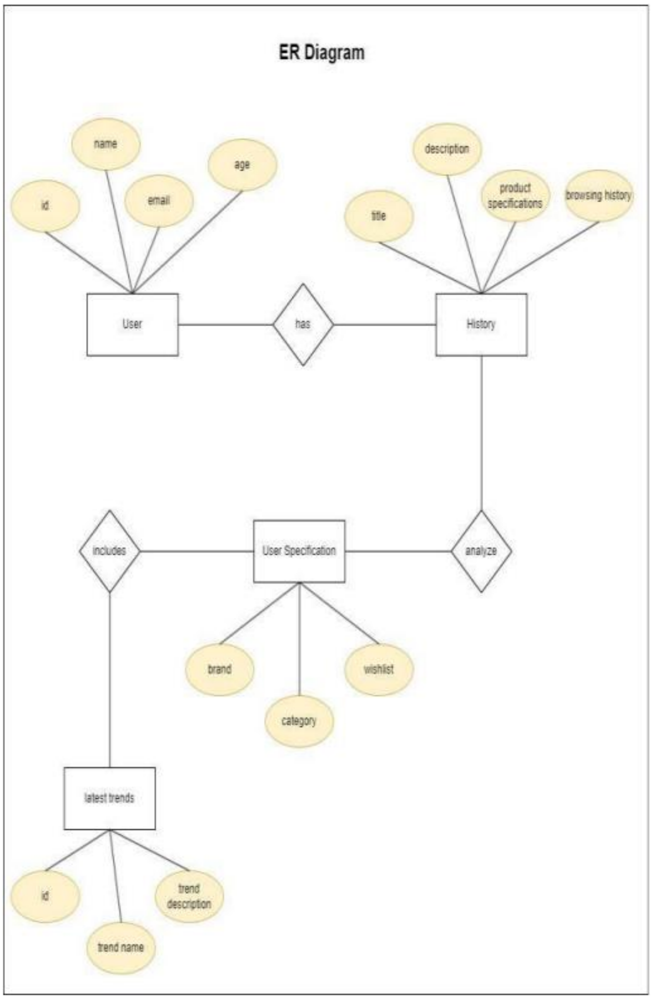

Fig 5 ER Diagram

V. CONCLUSION

The "Conversational Fashion Outfit Generator powered by GenAI" represents a remarkable fusion of artificial intelligence, fashion, and user interaction. Throughout this report, we have explored the methodology, features, user experience, impact on the fashion industry, challenges, and future improvements of this innovative technology. Additionally, we have highlighted case studies that exemplify its real-world applications and successes. The "Conversational Fashion Outfit Generator powered by GenAI" offers a transformative solution that not only simplifies fashion choices but also enhances the user's overall shopping and styling experience. It brings a new dimension to fashion by providing real-time, personalized recommendations and insights that resonate with individual preferences, styles, and occasions. Its impact on the fashion industry is far-reaching, contributing to sustainability efforts, boosting consumer engagement, and reshaping business strategies.

The development of this technology holds the potential to revolutionize how individuals engage with fashion and how fashion businesses operate. It represents a dynamic and promising intersection of AI and fashion, underlining the exciting possibilities and future advancements that lie ahead

ACKNOWLEDGEMENT

This paper was partially supported by Dr. D Y Patil Institute of Engineering Management and Research, Akurdi. This paper was created under the guidance of Dr. Amol Dhakne.

REFERNCES

[1]. Adding Conditional Control to Text-to-Image Diffusion Models - Lvmin Zhang, Anyi Rao, and Maneesh Agrawala Stanford University, arXiv:2302. 05543v2 [cs.CV] 2 Sep 2023
[2]. Taming the Power of Diffusion Models for High-Quality Virtual Try-On with Appearance Flow, arXiv:2308.06101v1 [cs.CV] 11 Aug 2023
[3]. UNER: Universal Named-Entity Recognition Framework, arXiv:2010.12406v1 [cs.CL] 23 Oct 2020
[4]. Sadia Afrin. Weight initialization in neural network, inspired by andrew ng, https://medium.com/ @safrin1128/weight initializatio n-in-neural-network-inspired-by-andrew-ng e0066dc4a566, 2020
[5]. Yuval Alaluf, Omer Tov, Ron Mokady, Rinon Gal, and Amit Bermano. Hyperstyle: Stylegan inversion with hypernetworks for real image editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 18511–18521, 2022
[6]. Omer Bar-Tal, Lior Yariv, Yaron Lipman, and Tali Dekel. Multidiffusion: Fusing diffusion paths for controlled image generation. arXiv preprint arXiv:2302.08113, 2023
[7]. Zhe Chen, Yuchen Duan, Wenhai Wang, Junjun He, Tong Lu, Jifeng Dai, Thiab Yu Qiao. Lub Zeem Muag Hlov Pauv Adapter Rau KV KV Evet Tuab. International Textiles Conference, 2023
[8]. James Lehtinen, Jacob Munk Lehberg, Jacob Munk Lehberg Hasselgren, Samuel Laine, Tero Karras, Micah Aittala and Timothy Aila. Noise2noise: Learning image reconstruction without clean data. Proceedings of the 2018 35th International Conference on Machine Learning
[9]. Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein genera tive adversarial networks. In ICML
[10]. Shuai Bai, Huiling Zhou, Zhikang Li, Chang Zhou, and Hongxia Yang. 2022. Single stage virtual try-on via deformable attention flows. In ECCV
[11]. Andrew Brock, Jeff Donahue, and Karen Simonyan. 2018. Large scale GAN training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096 (2018)
[12]. Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. NeurIPS (2017)
[13]. Sangyun Lee, Gyojung Gu, Sunghyun Park, Seunghwan Choi, and Jaegul Choo. 2022. High-Resolution Virtual Try-On with Misalignment and Occlusion-Handled Conditions. In ECCV
[14]. Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In CVPR
[15]. Bochao Wang, Huabin Zheng, Xiaodan Liang, Yimin Chen, Liang Lin, and Meng Yang. 2018. Toward characteristic-preserving image-based virtual try-on network. In ECCV
[16]. Na Zheng, Xuemeng Song, Zhaozheng Chen, Linmei Hu, Da Cao, and Liqiang Nie. 2019. Virtually trying on new clothing with arbitrary poses. In ACM MM
[17]. Akbik, A., Blythe, D., Vollgraf, R.: Contextual String Embeddings for Sequence Labeling. In: COLING 2018,27thInternationalConferenceon Computational Lin guistics. pp. 1638–1649 (2018)
[18]. De Araujo, P.H.L., de Campos, T.E., de Oliveira, R.R., Stauffer, M., Couto, S., Bermejo, P.: Lener-br: A dataset for named entity recognition in brazilian legal text. In: International Conference on Computational Processing of the Portuguese Language. pp. 313–323. Springer (2018)
[19]. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: Dbpedia: A nucleus for a web of open data. In: The semantic web, pp. 722–735. Springer (2007)
[20]. Honnibal, M., Montani, I.: spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing (2017), to ap pear
[21]. Kokkinakis, D., Niemi, J., Hardwick, S., Lind´en, K., Borin, L.: HFST-SweNER—A New NER Resource for Swedish. In: LREC. pp. 2537–2543 (2014)
[22]. Sang, E.F., De Meulder, F.: Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. arXiv preprint cs/0306050 (2003)