Flipkart GRID 5.0

(Conversational Fashion Outfit Generator powered by GenAI)

Attack Titans

Team Member Details:

Team Name	Attack Titans		
Institute Name	IIT (ISM) Dhanbad		
Team Members >	1(leader)	2	3
Name	Muskan Kharyal	Nishtha	Neha Singh
Batch	2021-2025	2021-2025	2021-2025

Flipkart GRiD 5.0

Fashino-GPT

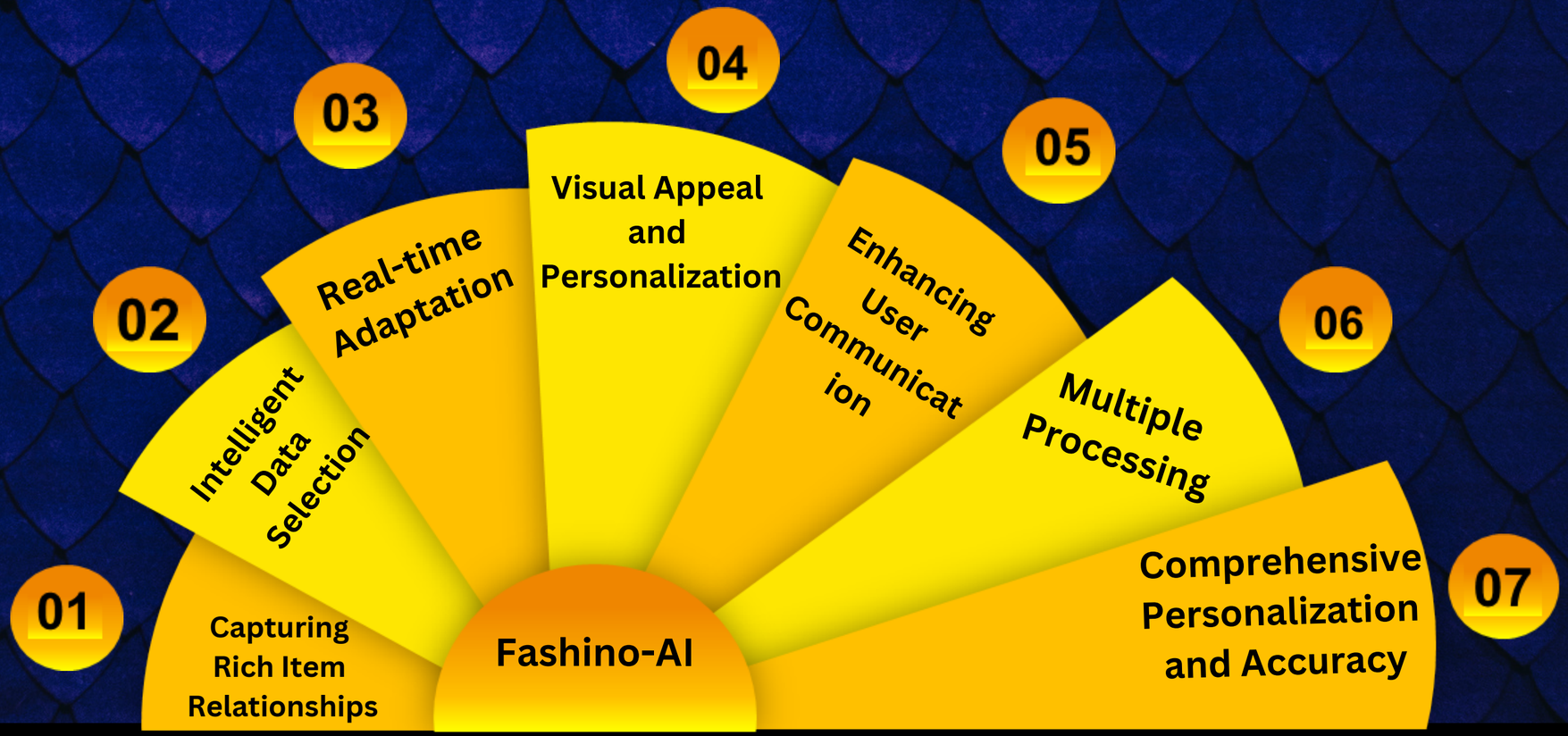
Conversational Fashion Outfit Generator powered by GenAI

Approach:

1. Extracting User Needs
2. Finding Similarity using NLP
3. Text Extraction & Model Training
4. Recommendation & optimization
5. Text Inversion

Product USP:

Fashino-AI is an intuitive heuristic-based solution that Captures Rich Item Relationships using node2vec making it less expensive than other approaches. Taking user preference as input we use graph active learning for intelligent data selection. It allows the system to find product similarities and optimize the learning process, focusing on informative data, leading to better results. For real-time adaptation, the fashio-ai is using stable diffusion for generating images with the help of input text by the user. Visual Appeal and Personalization are done using Image Retrieval and if the user is unsatisfied by the result, Text Inversion has been embedded. By generating images based on user preferences, the system is offering a personalized and visually appealing shopping experience, increasing user engagement. It is not only enhancing the accuracy of fashion recommender but also providing a well-rounded user experience that cis considering various dimensions of user preferences, interactions, and aesthetics.



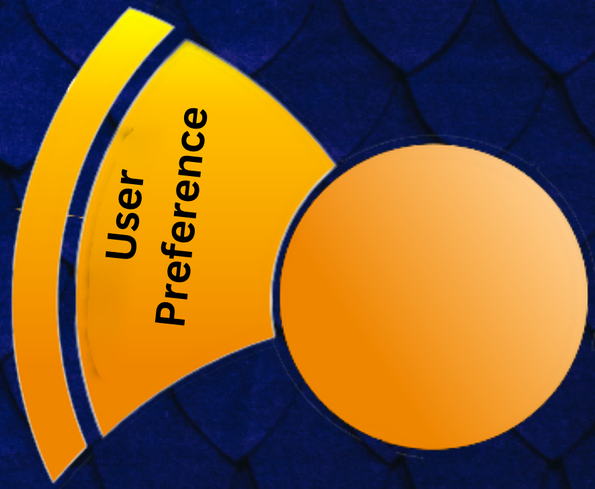
Proposed System Architecture

We take user preference using the node2vec algorithm, when the user gives their preference. Then product similarity is done using GAL. The GAL algorithm then gives its result. It is then followed by stable diffusion, here, the user gives a caption and AI generates an image accordingly. Image retrieval checks the stable diffusion generated image and dataset images in the AI system, giving a suitable image according to user preference. Lastly, if the user is not satisfied with the image generated by RS then textual inversion comes into play and then it suggests other suitable images.

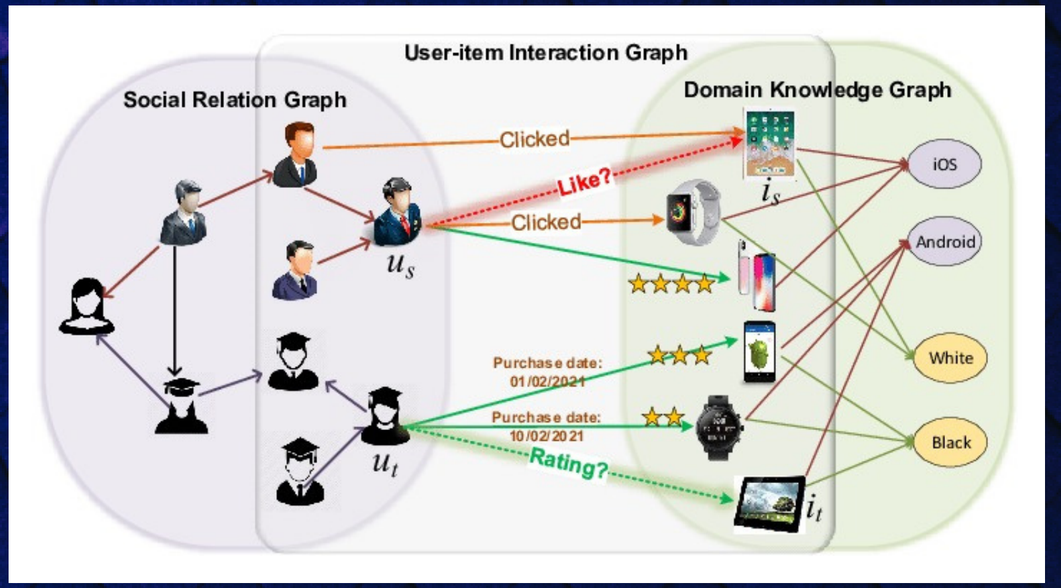
- Node2vec algorithms when we take user preference
- Graph active learning use for product similarity
- LLM is also used in Fashion RS
- Stable diffusion use for
- caption-image generation
- Image retrieval as shuffling a user preferred image
- Textual inversion as swapping of image if user doesn't like

Phase I:

- User interaction.
- Build graph and generate random walks.
- Node embedding and personalised recommendation
- User feeding and model iterations.



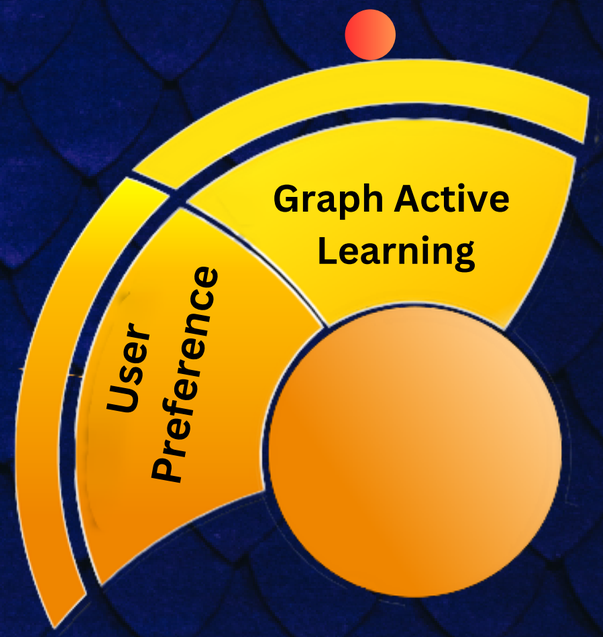
Node2vec algorithm in user preference has been given the task to find user preference.
It employs biassed random walks to generate node embeddings in a fashion item graph. It is capturing nodes' structural context, representing items as vectors. These embeddings, based on node sequences, enable fashion recommender systems to understand user preferences by measuring proximity and enhancing item recommendations through learned item associations, optimizing for personalized suggestions.


User-item Interaction Graph
Social Relation Graph
Domain Knowledge Graph
Clicked
Like?
Clicked
u_s
i_s
iOS
Android
White
Black
Purchase date:
01/02/2021
Purchase date:
10/02/2021
Rating?
u_t
i_t

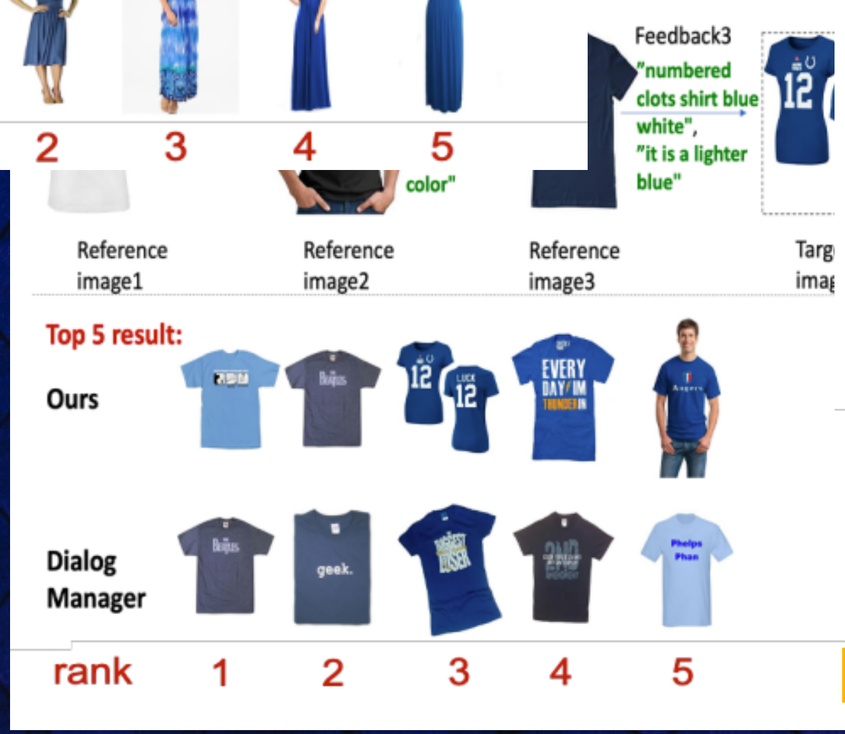
Phase II:

(Finding user product similarity using active learning)

- Initial data collection(image with description)
- Build similarity graph
- Select information samples and label selected samples
- Update ,evaluate and iterate model



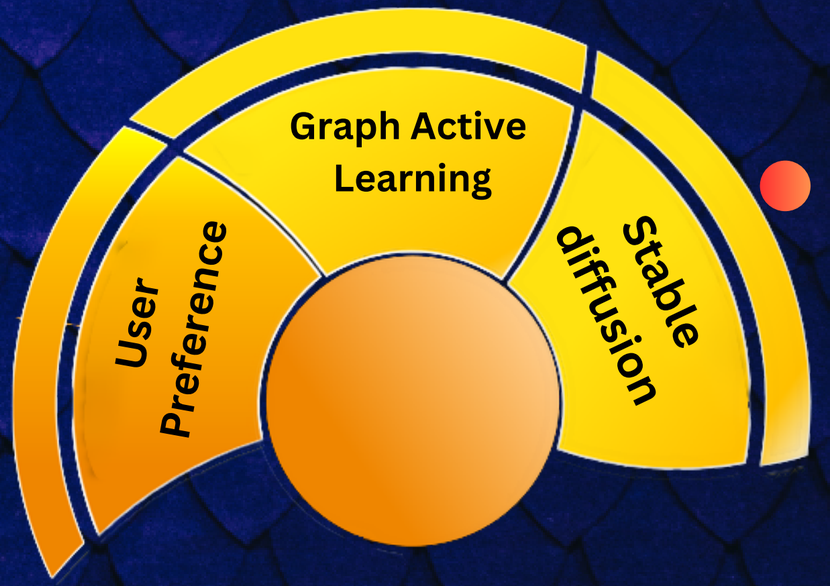
Graph Active Learning(GAL) optimises user-product similarity in fashion recommenders by iteratively selecting graph nodes for labelling. This refines embeddings and similarity estimation by leveraging labelled data effectively. By targeting relevant nodes, it enhances user-product similarity predictions, improving personalised fashion recommendations through informed sampling, efficiently utilising graph structure to enhance the system's understanding of user preferences and item relationships.



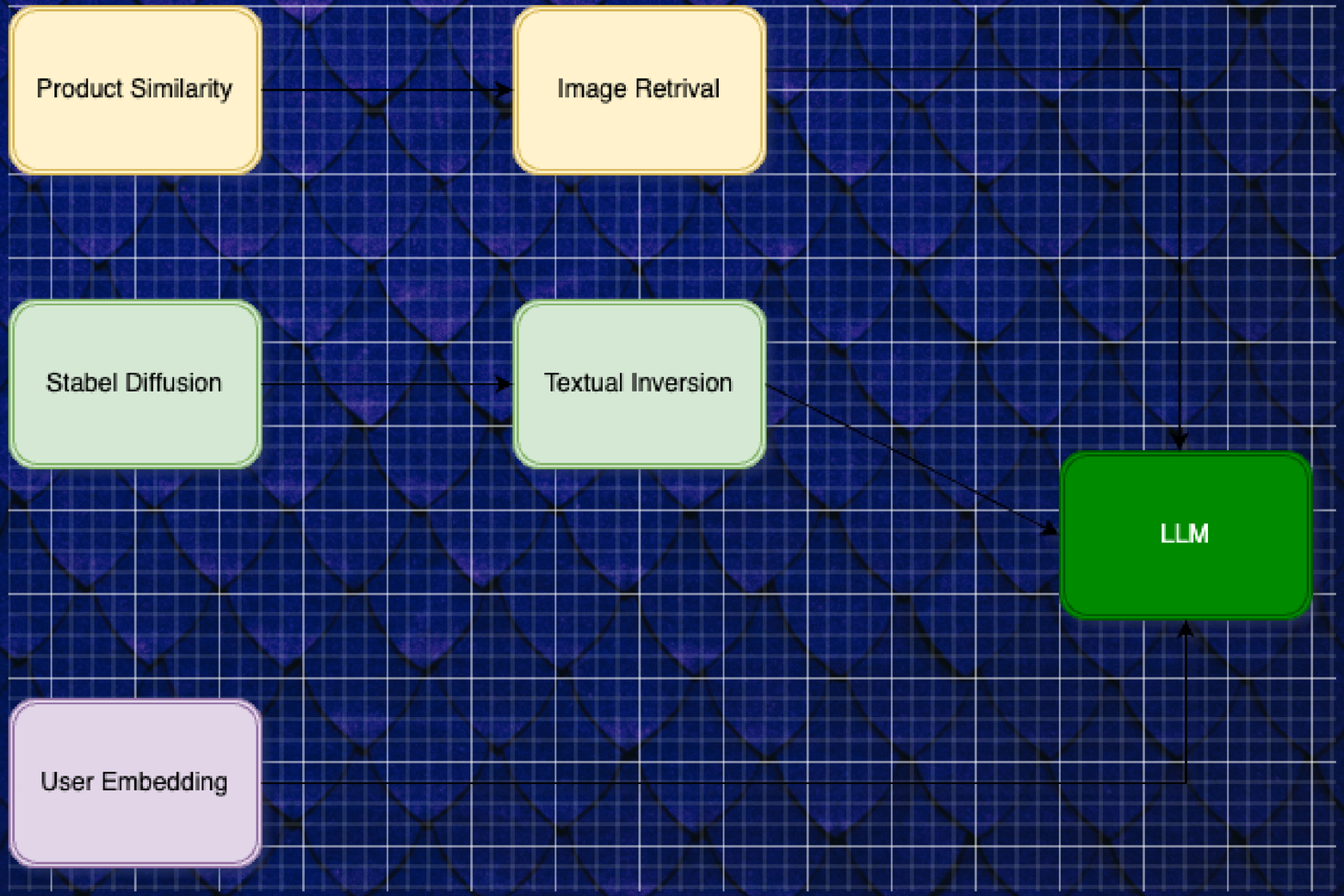
Examples of a product similarity

Phase III:

- Data processing
- Train stable diffusion model and train image generation model
- Generate image using captions
- Present enhanced recommendations to user and user feedback



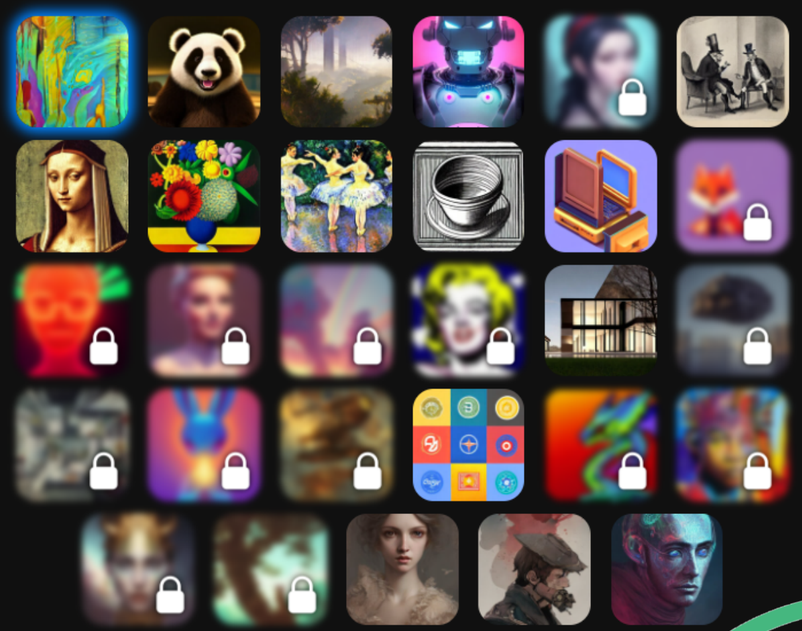
Stable diffusion model for text-to-image conversion has been considered here.
This model is employed in fashion recommendations by using graph-based propagation. It is capturing nuanced relationships among items using diffusion processes, generating stable node embeddings. By associating textual captions with image nodes, it is integrating captions as node attributes. The model refines embeddings iteratively, enabling improved image-caption associations. Fashion recommendations have been enhanced by aligning images and captions effectively within a graph-based framework.


Product Similarity
Image Retrival
Stabel Diffusion
Textual Inversion
LLM
User Embedding

Create an image from text prompt

diwali outfit on female dummy

Choose a style



Options

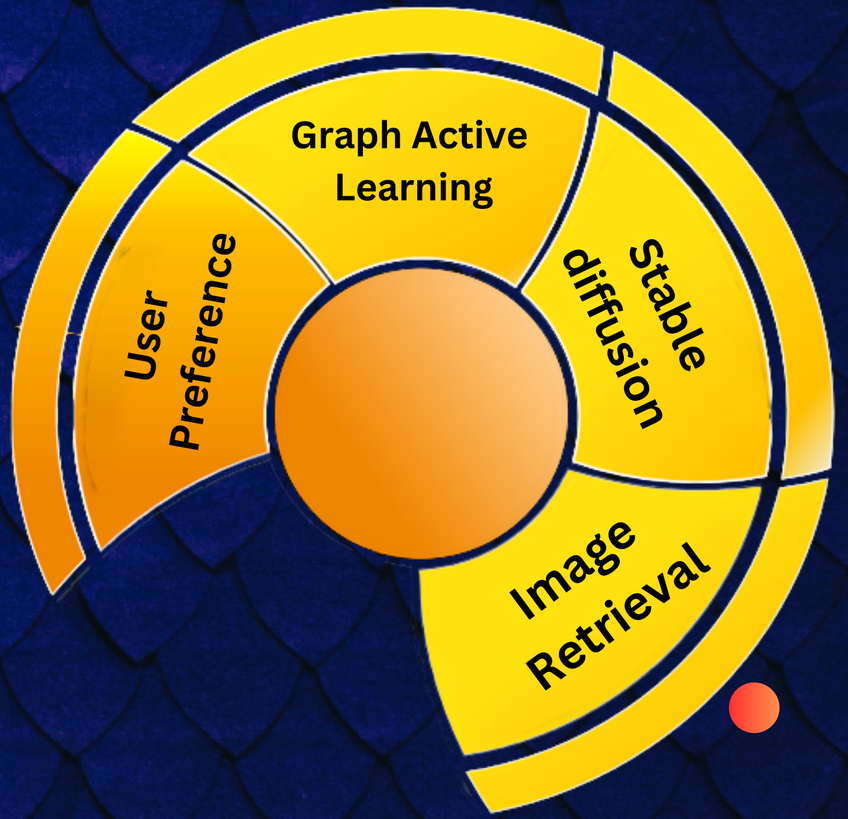
Generate

Download ⬇

Enhance ✦

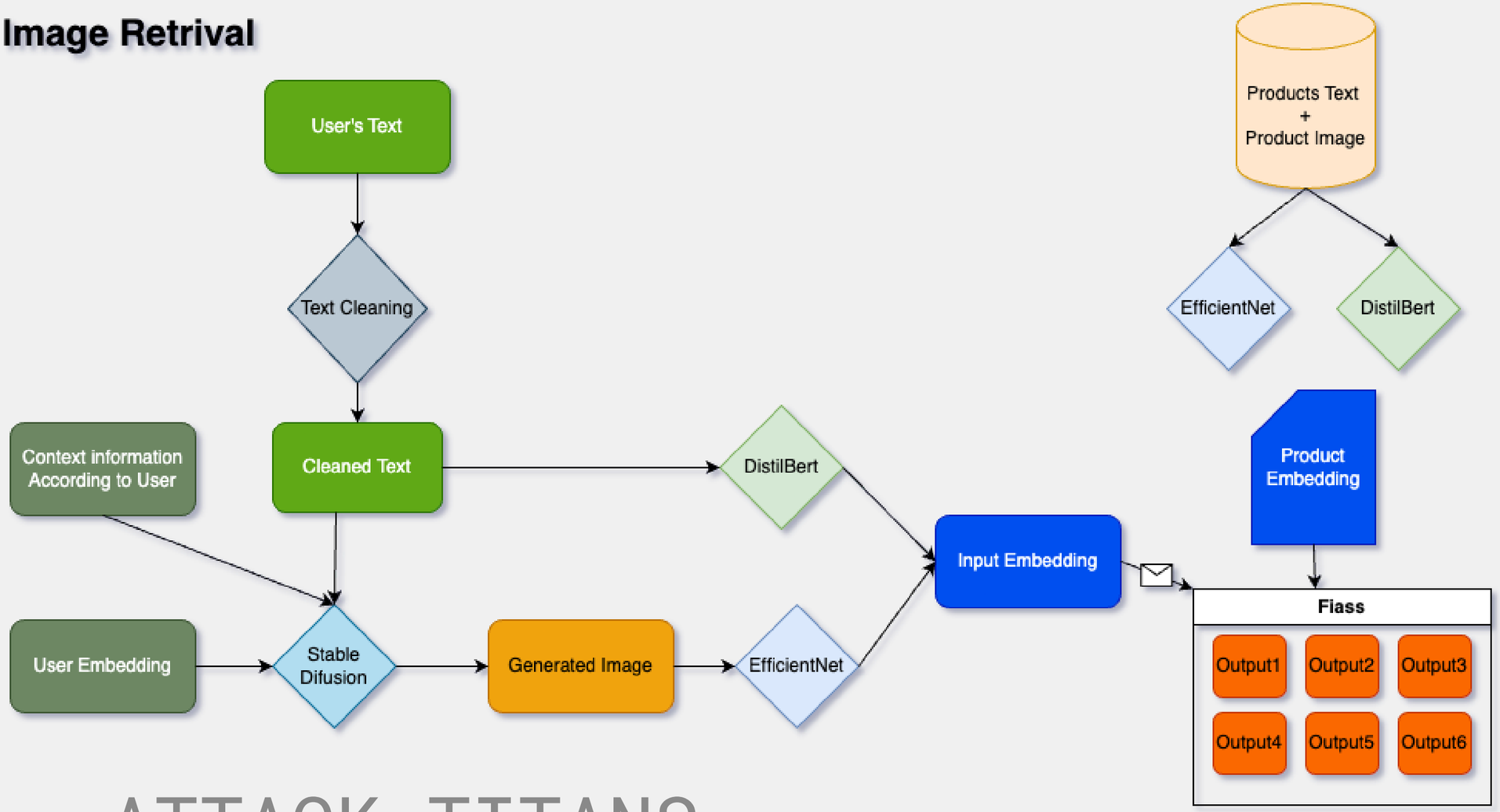
Phase IV:

- Query image submission
- Image retrieval and recommended products
- User feedback and updation accordingly
- System learning improvement and continuous user interaction



Post-generation from the stable diffusion model, image retrieval deploys feature extraction (e.g., CNNs, embeddings) for visual representation. It quantifies visual proximity using metrics (e.g., cosine similarity, Euclidean distance) to gauge similarity between generated and real images. By ranking real images based on these scores, it identifies tangible fashion items that closely align with generated visuals. This fine-tunes personalised fashion recommendations by translating the generated images into concrete products, augmenting user-specific preferences through precise visual matches.

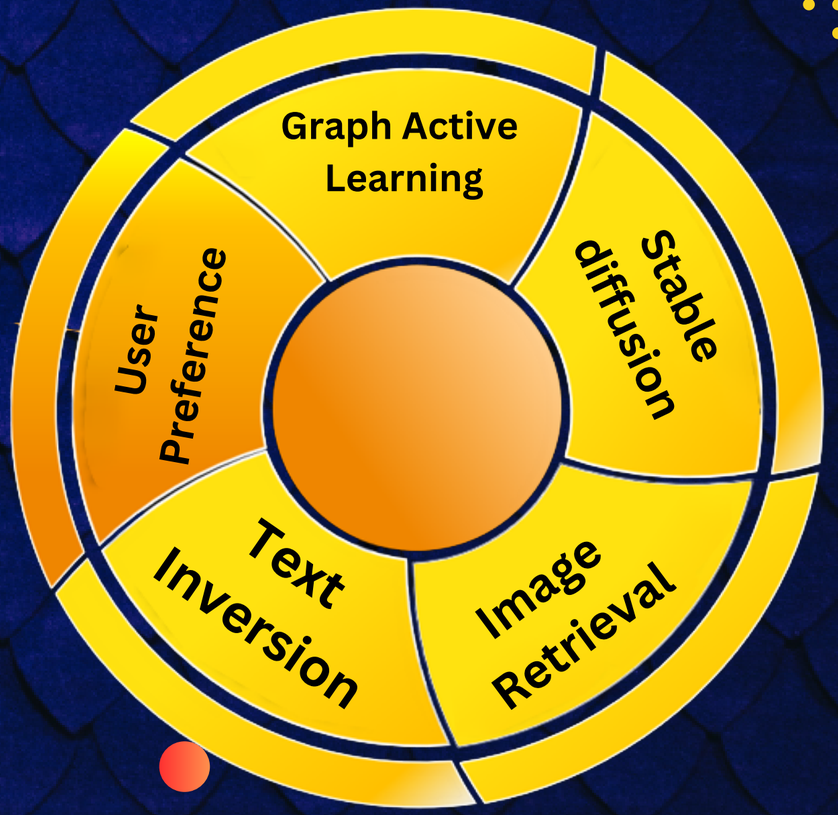
Image Retrival



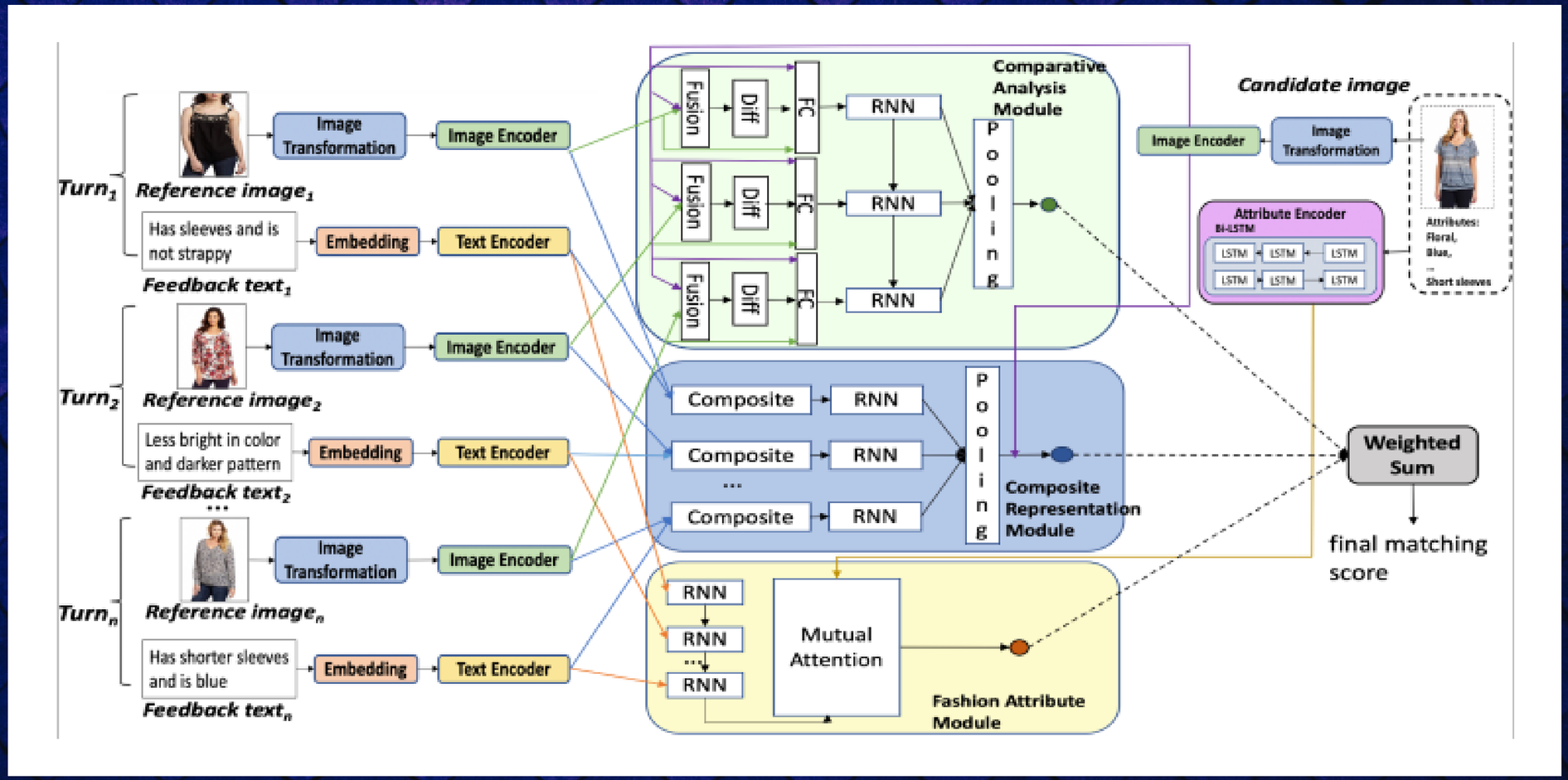
ATTACK TITANS

Phase V:

- Textual description input
- Textual inversion process and new result
- User feedback and update preferences based on feedback
- System improvement and continuous user interaction



If the user is not satisfied, Text Inversion will do so!
After image comparison through retrieval, textual inversion is employed. It reverse-engineers image embeddings to infer textual descriptions. Utilising techniques like neural networks, it decodes visual features into attributes (e.g., style, colour). These reconstructed descriptions enable dynamic text-based queries. By fusing textual attributes with user preferences, it enhances fashion recommendations. This process aligns visual information with textual descriptions, bridging the gap between visual and textual domains, and optimises personalised suggestions in the fashion recommender system through enriched attribute-based modelling.



Feedback System

Future Scope

01 AR/VR Integration

Implement augmented reality (AR) and virtual reality (VR) capabilities to enable users to virtually try on outfits and see how they look on themselves. This immersive experience could significantly enhance user engagement and satisfaction.

02 Collaborative Fashion Planning

Extend the system to support collaborative fashion planning, where users can invite friends or family members to join the conversation and offer input on outfit choices. This could mimic the experience of shopping together in a physical store.

03 Multi-lingual Support

Expand the system's capabilities to understand and respond in multiple languages, catering to a diverse user base and global audience.

04 Fashion Trend Forecasting

Utilize machine learning and data analysis to predict future fashion trends based on historical data and current influences.

Flipkart GRiD 5.0

Thank You