Master Thesis Exposé

A Picture is Worth a Thousand Words: Using AI Image Generation for Recommendations

Felix Timotheus Johannes Dietrich
Technical University of Munich (TUM)
felixtj.dietrich@tum.de

1. Introduction

With today's fragmented streaming service landscape, *e.g. Netflix, Disney+, Apple TV+*, there is a huge amount of content available to the user across multiple providers. Users browsing through the content of a streaming service are often overwhelmed by the sheer amount of content available. They end up spending a lot of time browsing through the content, but are often unable to find something they want to watch in a reasonable amount of time. This is known as choice overload which leads to the problem of decision paralysis [8].

The problem of decision paralysis is not limited to streaming services, but also occurs in other domains such as food, gifts, and travel. In these domains, the user is often presented with a large number of options, which leads to the user being unable to make a decision.

In this thesis, we propose a novel solution to the problem of decision paralysis by using AI image generation to gauge the user's preferences and provide them with a recommendation. Using AI image generation [13], we can generate images that represent potential preferences of the user and let the user make choices between these images. This method has the potential to capture the user's current mood and other subconscious preferences. Based on the user's choices, we can then provide them with a recommendation. A picture is worth a thousand words, and we can use this to our advantage.

2. Related Work

Recent advances in AI image generation [3, 5, 11, 13] have made it possible to create realistic images and art using deep neural networks. Popular text-to-image generation models include *DALL·E 2* [11] and *Stable Diffusion* [13] which allow users to generate images based on a description in natural language.

Recommendation systems [12] are used by many big companies such as Amazon [14], Netflix [2], and Youtube [1] to provide users with personalized recommendations and have a huge business value. There are many different approaches to recommendation systems, such as collaborative filtering [16], content-based filtering [10], and deep learning [17], to name a few.

Visual or image-based recommendation systems [6, 9] have been previously used in the context of fashion to recommend items based on the user's preferences and the style of the item.

To the best of our knowledge, there are no attempts to use AI image generation for guessing the user's preferences and providing them with a recommendation.

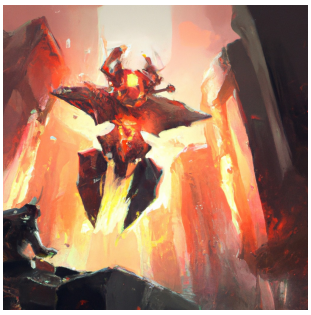
3. Proof of Concept

We have developed a proof of concept for our proposed solution during *hackaTUM 2022* [4] for the challenge provided by *StreamPicker* [15]. The proof of concept can be found at `zapped.ospatos.com` and extends on our hackathon submission at `hackatum22.ospatos.com`.

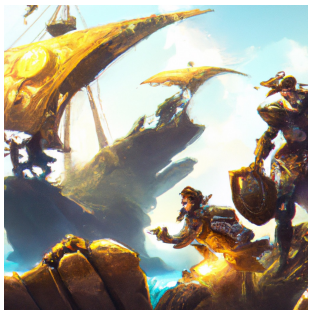
For the proof of concept, we pregenerated 1000 images using the *DALL·E 2* [11] text-to-image generation model. The images were generated using the following prompt pattern: *"Trending on artstation. Impressive digital art with excellent composition. genre1, genre2, genre3, keyword1, keyword2, keyword3"* filled with a popular movie's genres and keywords. Examples of the generated images can be seen in Figure 1. Out of the 1000 images, we select three times three random images and present them to the user to choose between. Based on the user's choices, we weight the genres and keywords associated with the chosen images and use them to generate a movie recommendation.

The results of the proof of concept are promising with mostly positive feedback from the users. Common feedback was that the images were visually appealing and that the users liked the idea of using AI image generation to capture their preferences and help them make a decision. On the other hand, users also pointed out that the images were

not always understandable and often too abstract, dramatic, mysterious, or dark (*cf*. Figure 1). A lot of things about this proof of concept can be improved, but we believe that the concept is promising and can have a positive impact on users' decision making.



(a) **Star Wars:** *Action, Science Fiction, Adventure, superhero, super power, space*



(b) **Pirates of the Caribbean - The Curse of the Black Pearl:** *Adventure, Action, Fantasy, pirate, gold, swashbuckler*

(c) **The Lord of the Rings - The Fellowship of the Ring:** *Fantasy, Adventure, Action, magic, battle, sword and sorcery*

(d) **Fight Club:** *Drama, Thriller, Comedy, dystopia, fight, dual identity*

Figure 1. **Example images generated by DALL·E 2.** The movie images are generated using the DALL·E 2 model [11] given prompts including the movie's genres and keywords.

4. Research Questions

1. How could a recommendation system using AI image generation work in detail?
2. Can we use machine learning techniques to progressively let the user visually choose concepts to build the movie they want to watch? (*e.g*. first choose how much romance, then how much action, then how much comedy, *etc*.)
3. How can we effectively control the image generation process to be visually appealing and understandable while still capturing the user's preferences?
4. How can the user interface be designed to effectively present the generated images and facilitate the decision-making process?
5. How can the use of AI image generation for personalized recommendations be evaluated and compared to traditional recommendation algorithms in terms of effectiveness and user satisfaction?

5. Method

In our research, we will use the diffusion-based model *Stable Diffusion* [13] for AI image generation, since it is a powerful and publicly available model that can generate high-quality images. We use movie data from the *TMDb* [7] movie database. We will also develop a web interface that allows users to interact with our movie recommendation system.

To capture user preferences and make recommendations, we will use a combination of techniques, including deep learning, deep generative networks, style transfer, fine-tuning, clustering, choice architecture, and other machine learning techniques. We will also conduct a user study to evaluate the effectiveness of our proposed approach compared to traditional methods for movie recommendations.

Overall, our methods and techniques will allow us to capture user preferences using AI image generation, and make accurate and personalized recommendations for movies.

6. Conclusion

The quality of AI image generation has improved significantly in recent years and has only begun to scratch the surface of its potential. In this thesis, we have proposed a novel application of AI image generation for using it to capture user preferences and make personalized recommendations for movies, illustrated it using our proof of concept. While our proof of concept has room for improvement, we are confident that its concept is promising and has the potential to improve users' decision making.

References

[1] Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*, pages 191–198, 2016. 1

[2] Carlos A Gomez-Uribe and Neil Hunt. The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems (TMIS)*, 6(4):1–19, 2015. 1

[3] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 1

[4] HackaTUM. Official hackathon of the Department of Computer Science at the Technical University of Munich. https://hack.tum.de/, 2022. 1

[5] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020. 1

[6] Qiang Liu, Shu Wu, and Liang Wang. Deepstyle: Learning user preferences for visual recommendation. In *Proceedings of the 40th international acm sigir conference on research and development in information retrieval*, pages 841–844, 2017. 1

[7] TiVo Platform Technologies LLC. The Movie Database (TMDB). http://www.themoviedb.org, 2022. 2

[8] Adriana Manolică, Andreea-Sînziana Guță, Teodora Roman, and Lorin Mircea Dragăn. Is consumer overchoice a reason for decision paralysis? *Sustainability*, 13(11):5920, 2021. 1

[9] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, pages 43–52, 2015. 1

[10] Michael J Pazzani and Daniel Billsus. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007. 1

[11] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022. 1, 2

[12] Paul Resnick and Hal R Varian. Recommender systems. *Communications of the ACM*, 40(3):56–58, 1997. 1

[13] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022. 1, 2

[14] Brent Smith and Greg Linden. Two decades of recommender systems at amazon. com. *Ieee internet computing*, 21(3):12–18, 2017. 1

[15] StreamPicker. Dein guide durch den streaming-dschungel! infos rund ums streamen von filmen und serien bei streaming-anbietern wie Netflix und Amazon Prime. hier findest du den Überblick! https://www.streampicker.de, 2022. 1

[16] Xiaoyuan Su and Taghi M Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009, 2009. 1

[17] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)*, 52(1):1–38, 2019. 1