Algorithmic Affordances: a technical report of a recommender based streaming service

Nadine Kanbier | 4283724

Utrecht University

Words: 1566

Abstract

Recommender systems have become an integral part of the movie streaming services. This paper will present the technical considerations behind a recommender-based movie streaming service (The Legal Pirate Bay), using the data-driven feedback loop as a framework. The target audience (aged between 18 and 44) is described, and their corresponding values of *autonomy*, *efficiency* and *diversity* are outlined. The metrics for the user input are enlisted, which are used for the data analysis. The three algorithms used for the analysis and recommendation are explained. Shortcomings and suggestions for potential improvements are discussed.

1. Introduction

Recommender systems have become an integral part of the movie streaming services. These systems help users decide what they could watch next, help explore a variety of options, help users compare those relevant options and they help users discover options and opportunities they might not themselves have imagined. While some of these technologies are deployed in explicit, interactive uses, other advances are employed behind the scenes, such as automatically filtering content based on relevance. While such attempts at personalization might be welcome when aligned with user's preferences, it can be the source of information hiding and actions at odds with user goals and expectations (Amershi et al., 2019). Recommender systems not only have an impact on users individually, but they also have an impact on society more broadly. After all, they shape user preferences and guide choices (Milano, Taddeo, & Floridi. 2020).

Because of this impact, there is a need for a more human-centric design of these recommender systems. Supported by the data-driven feedback loop tool (Figure 1), we designed a recommender based streaming service, named The Legal Pirate Bay.

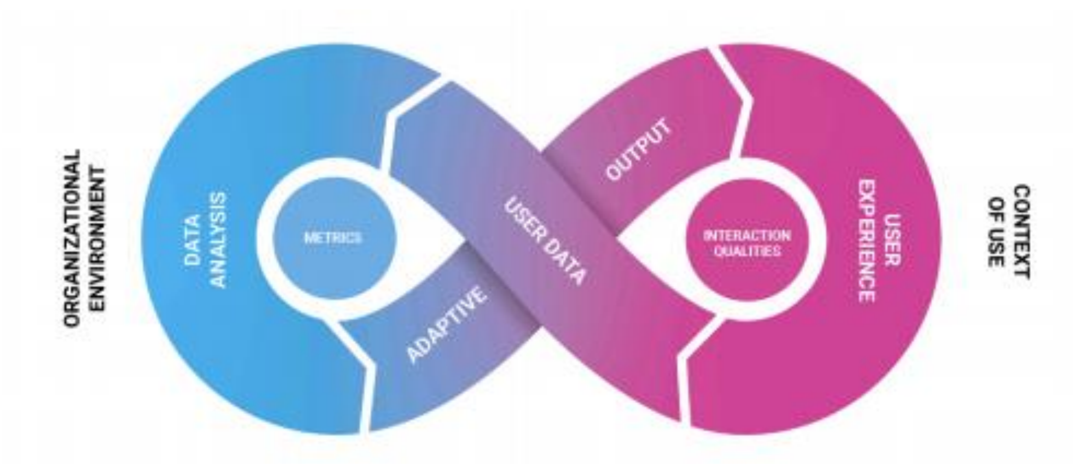


Figure 1. *Data-driven feedback loop by Smits et al. (2020).*

The data-driven feedback loop suggests an optimization process. On the user side of the loop, the focus is on qualities of the interaction (interface design). The choices on the user side of the loop are translated into metrics, to give reliable information to the organizational environment of the loop. A data-driven service can only be successful if there is a sound mapping between the interaction qualities on the one hand, and the metrics on the other hand (Smits et al., 2020). The data analysis on the organizational side of the loop creates an adaptive output that leads back to the users' side.

2. Target audience

The streaming service is targeted towards the largest group of streaming services: users aged 18-44 year old[1]. The main reason people use streaming services is entertainment, followed by informational and educational purposes (Lagger, Lux & Marques, 2017). We decided to focus on three values of the user: autonomy, efficiency and diversity. They will be discussed in the following section.

3. Values

Value 1: autonomy

People need to feel in control of their own behaviors and goals to feel self-determined (Deci & Ryan, 2008). Recommender systems often appear as a "black box", where users do not have insight into the logic of the system (Sinha & Swearingen, 2002). This black box nature of recommender systems prevents users from comprehending recommended results and can lead to the feeling of not being in control of your own behavior and goals (Herlocker, Konstan, & Riedl, 2000). As Helberger, Karpinnen & D'Acunto suggest in their 2018 paper on exposure diversity, it is possible to give back autonomy to users by providing the

[1] https://www.statista.com/statistics/742108/netflix-subscription-adults-usa-by-age/

possibility to choose between different recommendation logics. Taking this into account, together with guidelines 1, 2, 7, 11 and 17 of the Guidelines for Human-AI Interaction (see Appendix A, Amershi et al., 2019), we developed two separate pages on the streaming platform. One is non-personalized (Appendix B) and the other one is personalized (Appendix C). This clear distinction between non-personalized and personalized can potentially affect the user's mental model and change the way the user interacts and thinks about the workings of the technology. Furthermore, to adhere to guidelines 1, 2, and 11 of the Guidelines for Human-AI Interaction, we have developed a separate page that explains the workings of the recommender system in understandable language (Appendix D).

Value 2: efficiency

Today's overload on available movies and TV shows can be overwhelming. Recommender systems can confront this "information overload" problem. We use three types of recommender systems, so that the user can use the streaming service as efficient as possible. Explanations of the recommender systems and their algorithms can be found in section 5 of this paper.

Furthermore, we have created a shared watching experience, making the selection of movies to watch together easier and more efficient. It lists all the movies you both have not watched yet and it recommends you the movies you probably both would like (Appendix E).

Value 3: diversity

Algorithmic filtering and adaption of content to personal preferences are often associated with a decrease of information diversity (Helberger, Karpinnen & D'Acunto, 2018). However, recommender systems can also be designed to complement a personal media diet or expose users to different viewpoints. Thus, it is possible to not only give users information that they have been looking for, consider relevant to them or are most likely to watch (Nguyen et al., 2014).

Platforms such as Netflix have a so-called “long-tail”. Most existing recommender systems, especially collaborative filter-based methods, cannot recommend tail movies due to the data sparsity issue. Recommending popular products is easier yet more trivial (Yin, Cui, Yao, & Chen, 2012). To counteract the cons of the “long-tail” phenomenon, we designed a random selection in our streaming platform (Appendix B). This counteracts the triviality of only recommending popular movies and gives the user a more diverse offer.

4. Metrics

We have designed multiple metrics in our streaming platform in order to create an adaptive output. These are as follows:

- Rating the user gives to the movie (scale of 1 to 5)
- Number of times a movie is being found/seen
- Duration of watching the movie

These metrics (user input) are used for the data analysis on the organizational side of the feedback loop. After analyzing, these metrics will create an adaptive output (e.g. recommendations). Following section will go into the used algorithms and mention the used metrics.

5. Description of the algorithms

We have used three types of recommender systems in our video streaming service: a simple recommender, a content-based recommender and a collaborative filtering recommender.

1. Simple recommender

The first recommender, a simple recommender, is a basic system that recommends top items based on a certain metric or score. In our streaming service, this would be the rating of the movie. It would give the user the best-rated movies. However, using rating as a metric has a few shortcomings. It does not take the popularity of the movie in consideration. Thus, this metric tends to favor movies with a smaller number of voters. Taking these shortcomings into account, we have used a weighted rating formula as a metric. This is represented as follows:

$$\mathrm{W}\mathit{eightedRating}(\mathbf{WR}) = \left(\frac{\mathbf{v}}{\mathbf{v}+\mathbf{m}} \cdot \mathbf{R}\right) + \left(\frac{\mathbf{m}}{\mathbf{v}+\mathbf{m}} \cdot \mathbf{C}\right)$$

v is the number of votes for the movie

m is the minimum votes required

R is the average rating of the movie

C is the mean vote across the whole report

In our recommender system, we have used a cutoff of 90% (it must have more votes than at least 90% of the movies on the list). Applying this function to our dataset gave us our top-rated movies. This recommender system is used in the non-personalized section of the streaming service.

Value: efficiency.

Used metric: rating of the movie.

2. Content-based recommender

The second type of recommender we have used is a content-based recommender system. This type of system recommends similar items based on another item. Because of this,

this recommender system is used in the personalized page of the streaming service. It gives movies based on movies the user already watched.

To compute this in the back end, we have computed pairwise cosine similarity scores for all movies based on their plot descriptions. We started with vectorizing the descriptions of the movies and computing the Term Freqency-Inverse Document Frequency (TF-IDF) vectors for each of the movies. After this, we chose cosine similarity to calculate a numeric quantity that denotes the similarity between movies. Cosine similarity has proven to be a robust metric for scoring the similarity between to strings (Tata & Patel, 2007). Mathematically, it is defined as follows:

$$\cos(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}^{\mathsf{T}}}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} = \frac{\sum_{i=1}^{n} \mathbf{x}_i \cdot \mathbf{y}^{\mathsf{T}}{}_i}{\sqrt{\sum_{i=1}^{n} (\mathbf{x}_i)^2} \sqrt{\sum_{i=1}^{n} (\mathbf{y}_i)^2}}$$

This function gave us the possibility to recommend similar movies to the user, based on earlier activity.

Value: efficiency.

Used metric: Having watched more than half of a particular movie, not only having clicked it.

3. *Collaborative filtering recommender*

   The third recommender we used is a collaborative filtering recommender. Collaborative filtering recommenders try to predict the rating or preference that a user would give an item based on past rating and preferences of other users. This means that this system is also used in the personalized page of the streaming service, and not in the non-personalized.

   Value: efficiency.

   Used metric: ratings.

6. Discussion

There are several shortcomings in this technical design.

First of all, as this is only a concept of the design, we would need more testing to see if the parameters, metrics and algorithmic affordances are connected to the three mentioned values. Furthermore, coupling between action and perception is not so simplistic: the focus on perception neglects cultural meaning and human learning. Continuous testing and improving of the design is necessary to achieve aforementioned values and goals.

Second, this recommender system does not take a cold start into account. This means that they cannot make effective recommendations for new users or for new items that have no explicit or implicit relevance indicators yet.

Third, there is always a trade-off between values. There are value tensions such as autonomy vs. efficiency (there is an extra click to the personalized page and a second extra click to the explanation, which is not efficient but creates a feeling of autonomy). Also, there is an engagement/diversity trade-off (Holz et al., 2020). These trade-offs will always be there.

Further improvement of the streaming service would be in the direction of applying more filters, so users can be even more selective and in control over their own recommendations.

References

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).

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian psychology/Psychologie canadienne*, *49*(3), 182.

Helberger, N., Karppinen, K., & D'acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, *21*(2), 191-207.

Herlocker, J. L., Konstan, J. A., & Riedl, J. (2000, December). Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work* (pp. 241-250).

Holtz, D., Carterette, B., Chandar, P., Nazari, Z., Cramer, H., & Aral, S. (2020, July). The Engagement-Diversity Connection: Evidence from a Field Experiment on Spotify. In *Proceedings of the 21st ACM Conference on Economics and Computation* (pp. 75-76).

Lagger, C., Lux, M., & Marques, O. (2017). What makes people watch online videos: An exploratory study. *Computers in Entertainment (CIE)*, *15*(2), 1-31.

Milano, S., Taddeo, M., & Floridi, L. (2020). Recommender systems and their ethical challenges. *AI & SOCIETY*, *35*(4), 957-967.

Nguyen, T., Hui, P.-M., Harper, F. M., Terveen, L., & Konstan, J. A. (2014). Exploring the filter bubble: the effect of using recommender systems on content diversity. Proceedings of the 23rd international conference on World Wide Web.

Sinha, R., & Swearingen, K. (2002, April). The role of transparency in recommender systems. In *CHI'02 extended abstracts on Human factors in computing systems* (pp. 830-831).

Smits, A., Van Turnhout, K., Hekman, E., & Nguyen, D. (2020). Data-driven design. In *DS 104: Proceedings of the 22nd International Conference on Engineering and Product Design Education (E&PDE 2020), VIA Design, VIA University in Herning, Denmark. 10th-11th September 2020*.

Tata, S., & Patel, J. M. (2007). Estimating the selectivity of tf-idf based cosine similarity predicates. *ACM Sigmod Record*, *36*(2), 7-12.

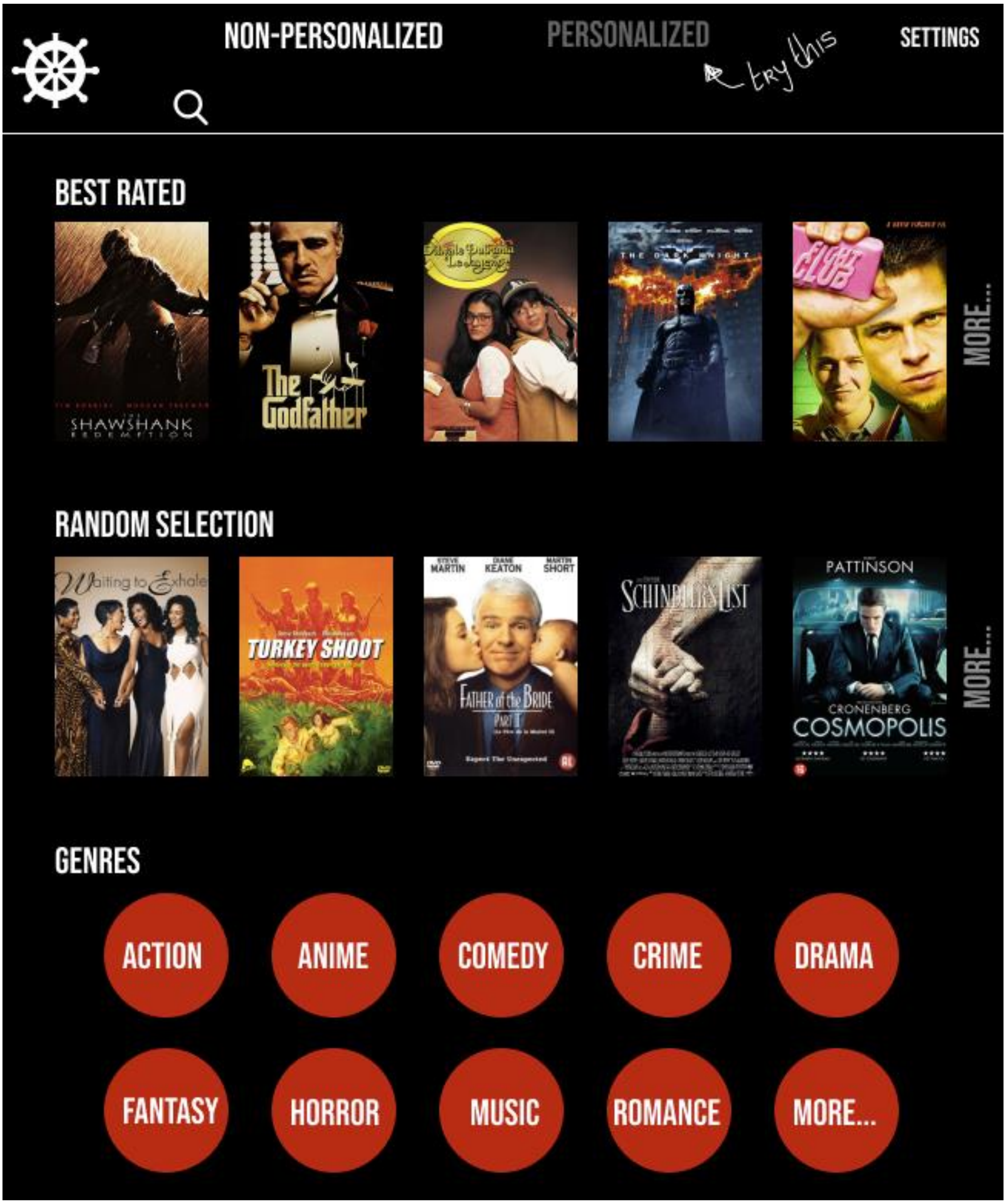
Yin, H., Cui, B., Li, J., Yao, J., & Chen, C. (2012). Challenging the long tail recommendation. *arXiv preprint arXiv:1205.6700*.

Appendix

Appendix A. Guidelines for Human-AI Interactions.

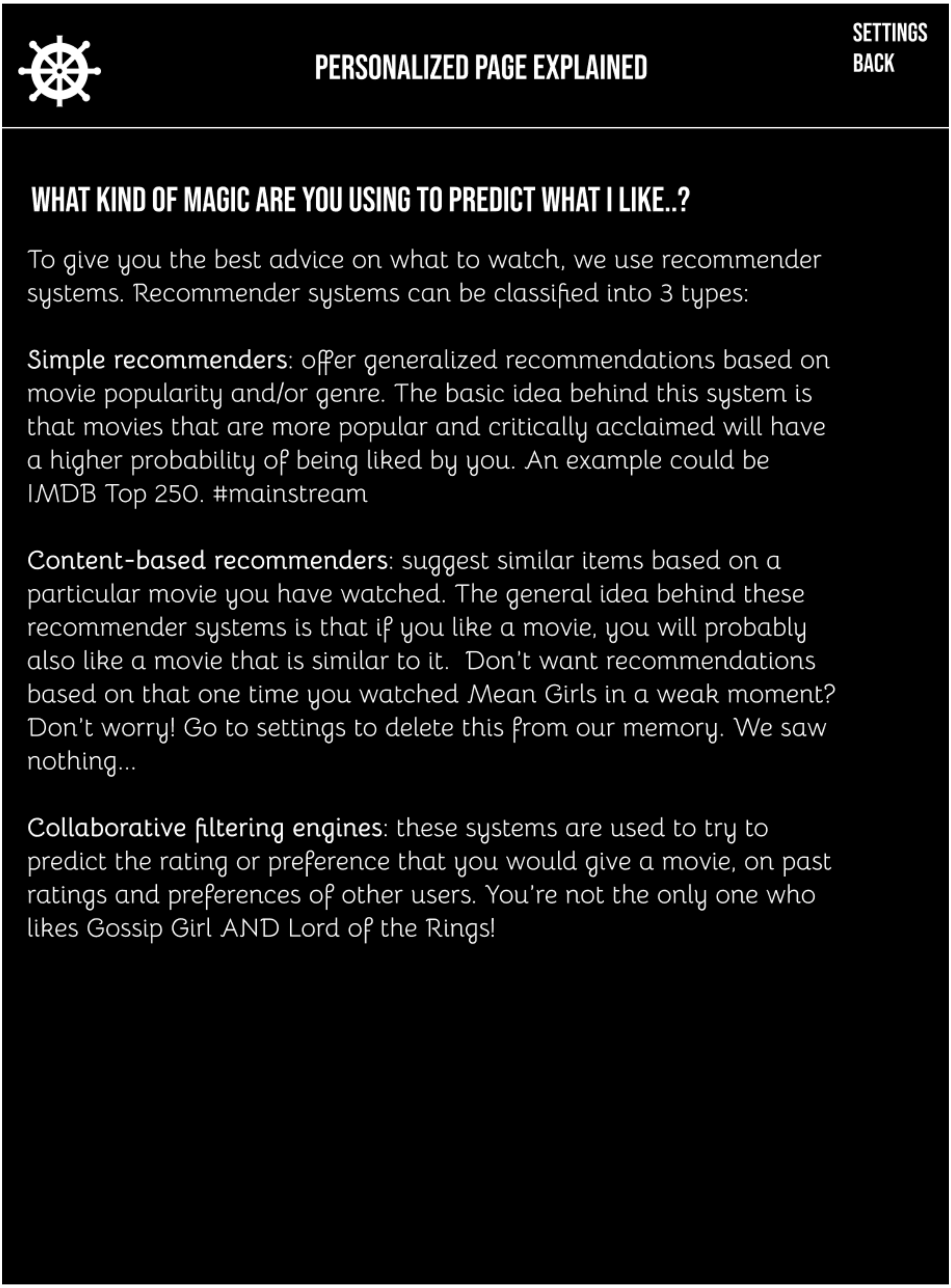
	AI Design Guidelines
G1	**Make clear what the system can do.** Help the user understand what the AI system is capable of doing.
G2	**Make clear how well the system can do what it can do.** Help the user understand how often the AI system may make mistakes.
G3	**Time services based on context.** Time when to act or interrupt based on the user's current task and environment
G4	**Show contextually relevant information**. Display information relevant to the user's current task and environment
G5	**Match relevant social norms.** Ensure the experience is delivered in a way that users would expect, given their social and cultural context.
G6	**Mitigate social biases**. Ensure the AI system's language and behaviors do not reinforce undesirable and unfair stereotypes and biases.
G7	**Support efficient invocation**. Make it easy to invoke or request the AI system's services when needed.
G8	**Support efficient dismissal**. Make it easy to dismiss or ignore undesired AI system services.
G9	**Support efficient correction**. Make it easy to edit, refine, or recover when the AI system is wrong.
G10	**Scope services when in doubt**. Engage in disambiguation or gracefully degrade the AI system's services when uncertain about a user's goals.
G11	**Make clear why the system did what it did**. Enable the user to access an explanation of why the AI system behaved as it did.
G12	**Remember recent interactions**. Maintain short term memory and allow the user to make efcient references to that memory.
G13	**Learn from user behavior.** Personalize the user's experience by learning from their actions over time.
G14	**Update and adapt cautiously.** Limit disruptive changes when updating and adapting the AI system's behaviors.
G15	**Encourage granular feedback.** Enable the user to provide feedback indicating their preferences during regular interaction with the AI system.
G16	**Convey the consequences of user actions.** Immediately update or convey how user actions will impact future behaviors of the AI system.
G17	**Provide global controls.** Allow the user to globally customize what the AI system monitors and how it behaves.
G18	**Notify users about changes**. Inform the user when the AI system adds or updates its capabilities.

Appendix B. Non-personalized page of the streaming service.



Appendix C. Personalized page of the streaming service.

Appendix D. Explanation page of the streaming service.



Appendix E. Shared page of the streaming service.