Recommendation Systems and Machine Learning: Mapping the User Experience

Luiz Agner [1], Barbara Necyk [2] and Adriano Renzi [3]

[1] Faculdades Integradas Helio Alonso; IBGE
Rio de Janeiro, RJ, Brazil
luizagner@gmail.com
[2] Universidade do Estado do Rio de Janeiro/ESDI; PUC-Rio
07barbara@gmail.com
[3] Universidade Federal Fluminense/UFF
adrianorenzi@gmail.com

Abstract. Human-algorithm interaction emerges as the new frontier of studies involving interaction design and interface ergonomics. This paper aims to discuss the effectiveness and communicability of streaming media recommendation systems, based on machine learning, from users' mental model point of view. We examined the content consumption practices on the Netflix platform, identifying some sensitive aspects of the interaction with recommendation algorithms. One-on-one semi-structured interviews were applied to a sample of students from three different universities in Rio de Janeiro, Brazil. We realised that interviewees had not correctly understood how the system works and have not formed an adequate mental model about tracked data and how it is processed to create personalised lists. Another issue concerns data privacy: Users have revealed a suspicion concerning algorithms and what might happen to usage data, not only in the Netflix platform but also in other services that use algorithm-based recommendation. Interviewees' responses suggested that there may be communication failures, and UX designers should strive to make it visible how the system tracks and processes user interaction data.

Keywords: User Experience, UX Design, Recommendation Systems

1 Introduction

This paper aims to study and discuss the effectiveness and communicability of entertainment content recommendation systems, based on machine learning, from users' mental model point of view. An exploratory interview research method was used to map how young college students in Rio de Janeiro, Brazil, interact with machine learning-based systems in entertainment applications such as Netflix.

Nowadays, there is a rush for applying artificial intelligence (AI) among major Internet companies, and interest grows in the same proportion as research investments.

Machine learning has emerged as a topic of great importance within the field of artificial intelligence. Bengio [1] cites artificial intelligence application's disruptive examples: industrial production, medicine, transportation, agriculture, personal assistants, translations, voice recognition, facial recognition, among other areas. Industry often applies AI as recommendation systems that feature personalised entertainment and information content for the user's interaction.

What is possible to know about the experience quality based on these interactions? Are users able to formulate a mental model on how recommendation systems work? Are users satisfied with custom recommendations? What is the mental model built by users regarding machine learning algorithms? Do users understand how machine learning operates and can interact with these algorithms to improve their outputs and better match them to their goals?

In this research, we chose to focus on interaction with the Netflix platform, studying the user experience on its recommendation algorithms. The reason for that choice is that Netflix is the most popular streaming service in the Brazilian market with broad dissemination among young people and is a reference concerning recommendation systems.

2 A Brief Introduction to Recommendation Systems

Nowadays, due to the Internet and information technology, people are facing the problem of enormous information overload and an ever-increasing amount of options to choose from, concerning several aspects of our lives. From choices for videos, music, books, travel itineraries, and restaurants, to health security, courses, or dating partners, the act of deciding is increasingly hard. Frequently, making a choice is a significant, challenging, and complicated task.

For instance, according to Pandey [2], the less popular products are abundant, yet they are not found easily in physical stores and are only available online, constituting what can be called a Long Tail [3]. Even these products may be adequate, but finding them on a website often becomes an arduous task. Therefore certain filters are important.

Machines are much more capable than humans in managing an extensive set of data and in learning from them. Thus, we can observe that computer programs, known as recommendation systems, may have a lot to assist us. They can be supporting the task of quickly establishing data generalisations and helping our decision-making processes, as Gomez-Uribe and Hunt [4] have suggested.

The so-called recommender systems (or recommendation engines) are programs that belong to the category of information filtering systems, and seek to predict which classification a user will give to an item of information, based on calculations and statistical inferences. According to Rocca [5], in a very generic way, recommendation systems can be defined as “algorithms designed to suggest items relevant to the user.” Those items can be movies to watch, texts to read, products to buy, or anything else depending on specific industries. According to Pandey [2], they aim to solve two types of problems: prediction (data used to predict the evaluation a user will give to an item he has not interacted with) and ranking (defining a finite list of items to be presented to the system's user).

For Rocca [5], recommendation systems now play an increasingly central role in our lives. From e-commerce (suggesting to buyers articles that might interest them) to digital marketing (suggesting to users the right ad that can generate the highest click-through rate), recommendation systems have now become inevitable in online journeys.

These systems are increasingly employed in commercial services that people use on a day to day basis, such as Netflix, YouTube and Spotify playlist generators, Amazon product offerings, or friend recommendations or social media content like Facebook, Linkedin and Twitter. According to Pandey [2], IMDB, Trip Advisor and Google News are other examples of platforms that continually send personalised suggestions and recommendations to their users.

The purposes of recommendation systems, according to Aggarwal *apud* Pandey [2], are:

(i) Relevance - The items to be purchased should interest users;

(ii) Novelty - This makes more sense to recommend items that users do not already know;

(iii) Serendipity - Recommend unexpected or surprising items;

(iv) Diversity - Internal variability in a list of items.

Recommendation systems often use the collaborative filtering approach or content-based filtering, as well as other approaches. Collaborative filtering builds a model based on past user behaviour (purchased or selected items, or ratings given to items), as well as similar decisions made by other users. This model can point out which items the user may be interested in. Content-based filtering approaches use discrete and pre-identified characteristics of each item to recommend new items with similar properties. The most widely used recommendation systems, on the market today, combine one or more approaches into one hybrid system.

2.1 Collaborative Filtering

Collaborative filtering can recommend complex items (e.g. series or movies) without presupposing an understanding of the content. That is because it is based solely on the - often questionable - assumption that people who have agreed in the past will also agree in the future, and people will like items similar to what they have liked in the past.

Rocca [5] points out that collaborative methods for recommendation systems are methods based only on past interactions recorded among users and items to produce new recommendations. These interactions are stored in the so-called "item-user interaction matrix."

The main advantage of collaborative approaches is that they do not require information about users or items and can, therefore, be used in a wide variety of situations or contexts. Also, the more the users interact with the items, the more accurate the recommendations become. That is, for a fixed set of users and items, interactions recorded over time generate new data and make the system increasingly effective, said Rocca

[5]. However, we should note that these collaborative filtering approaches often suffer from problems such as cold starting, which we will discuss later.

Furthermore, we should consider that the data collected to model user behaviour can be either explicit or implicit. Explicit forms of data collection include: search terms, asking a user to sort items or creating a wish list. Implicit forms of data collection include: observing browsing times, registering purchased products, viewing product lists, or analysing the social network. According to Pandey [2], implicitly collected data are derived from user interactions with the system and interpreted as indicators of interest or disinterest. With a broad set of tracked user data, it is possible to identify clusters representing user communities with very similar tastes to observe their collective behaviour (Kathayat) [6].

There are two subcategories within collaborative approaches: the so-called memory-based approaches and model-based approaches. The memory-based approaches assume that there is no model and are based only on records of observed interactions between close neighbours (for example, they search for users that are closest to a user of interest and find the most popular items among those neighbours). Model-based collaborative approaches assume that there is an underlying model capable of explaining interactions and attempt to identify it from a matrix of interactions to generate their predictions [5]. According to Pandey [2], these approaches employ machine learning methods.

2.2 Content-Based Filtering

According to Rocca [5], while collaborative methods depend only on item-user interactions, content-based approaches need additional information about users and items. They need a description of each item as well as a profile of user preferences. In this case, keywords are entered to describe the items and a user profile created. If we consider the example of a movie recommendation system, this additional information may be, for example, age, gender, profession or any other personal user data, as well as genre, main actors, duration, country of origin, awards or other features for movies.

The concept of content-based methods is based on building a model that explains the observed interactions among items and users. If the system can get this model, it is easy to make new predictions for a user: just have to check their profile (age, gender, etc.) and, based on that information, select relevant movies to suggest them.

Content-based recommendation systems may also ask users to write a review or some feedback about catalogue items. Based on what the user has liked in the past or is currently looking for, the algorithm finds and presents new similar items. According to Pandey [2], the disadvantage of these methods is their inability to learn from interactions so that there is no improvement over time.

2.3 Hybrid Recommendation Systems

Pandey [2] points out that both collaborative and content-based approaches have strengths and weaknesses. Therefore, most of the recommendation systems we currently use take a hybrid approach, merging collaborative filtering solutions with content-based approach methods, as well as other approaches. Approach hybridisation can be achieved by forecasting each method separately and then combining them, adding features from one approach to another or even unifying them into a single model.

Pandey [2] argues that Netflix is an excellent example of hybrid recommendation systems because its recommendations are based not only on browsing and consumption habits (collaborative approach) but also on similar feature videos (content-based). In order to do so, the company employs several algorithms, as we will see later.

2.4 Success Factors for Recommendation Systems

Nowadays, among the critical success factors of recommendation systems, one must consider points that go beyond the pure mathematical precision of recommendations, since the quality of the experience and user satisfaction involve many other equally relevant aspects. Some of the factors that impact the user experience are highlighted below:

(i) Diversity - Diverse recommendation lists tend to generate greater user satisfaction, so it is essential to recommend items that fit the users well but are not overly similar to each other, which avoids locking them into an “information bubble”;

(ii) Privacy - Creating user profiles using collaborative filtering can be problematic from a privacy point of view, and it is crucial to protect sensitive user data. European countries, as well as Brazil, have passed recent laws aimed at protecting their citizens’ data;

(iii) Demographics - Demographic aspects and other characteristics of users may become very relevant in the evaluation of recommendation systems. Beel, Langer, Nuenberger and Genzmehr [7] pointed out that maturer users tend to be interested in clicking on system recommendations more often than younger users;

(iv) Persistence - Resubmitting recommendations may be an appropriate strategy, as users tend to ignore items that are first presented to them due to a lack of time;

(v) “Rich gets richer” effect - According to Gomez-Uribe and Hunt [4], these approaches are still unable to deal with the strong bias caused by positive feedback, in which items that users are most involved with are recommended for many other members, leading to even greater involvement. The problem is especially real for collaborative memory-based algorithms (Rocca) [5], and tends to create an “information confinement area”;

(vi) Explainability - If users do not understand why a particular item was presented to them, they tend to lose confidence in the recommendation system, so they must be able to build a mental model of the system (Rocca) [5];

(vii) Response Time - Complexity poses a problem: Generating a new recommendation list can be time-consuming for recommendation systems that handle a massive amount of data for millions of users and millions of items.

(viii) Cold Start - This is a recognised condition that negatively impacts novice users’ experience in interacting with collaborative approach algorithms since, for these users, there is not enough data to generate accurate recommendations.

(ix) Serendipity - This feature refers to the system’s ability to recommend unusual items and surprise the user, not only give him miscellaneous items or novelties, as Pandey [2] noted. Consequently, Serendipity is distinct from simple Diversity.

Due to the listed circumstances, user satisfaction studies are the order of the day. In addition to the standard performance metrics captured in online reviews (A / B testing) and offline reviews, often criticised today, it is also essential to conduct qualitative studies and tests. They must be focused on user experience (UX) evaluation and describe her interaction with recommendation algorithms, such as the studies conducted by Budiu [8] and Harley [9] [10], which we will summarise later.

3 UX and Machine Learning Systems

According to Lovejoy and Holbrook [11], machine learning (ML) is the science of helping computers discover patterns and relationships in data instead of being programmed manually. Today, ML capabilities are becoming increasingly accessible to developers, and there is an impulse to be added to services to enhance the user experience. It's a powerful tool for creating personalised, engaging user experiences and is already directing everything from Netflix recommendations to autonomous cars.

As the authors said, as more user experiences incorporate ML, it is clear that UX designers still have a lot to learn about how to make users feel in control of the technology and not the other way around.

It is known that the most valuable ML systems evolve in connection with users' mental models. When people interact with these systems, they influence and adjust the kind of effects they will see in the future. These adjustments, in turn, will change the way users interact with the system, the models, and so on in a feedback cycle.

However, this can result in "conspiracy theories," in which people create incorrect or incomplete mental models of a system and try to manipulate the results according to imaginary rules. It is necessary to guide users to build accurate mental models that encourage them to give feedback. This fact will be mutually beneficial to them and the model (Lovejoy and Holbrook) [11].

In a study focused on users of ML-based systems, Budiu [8] inferred that people could not make the interfaces do what they want because they form weak mental models. Usability research has observed that these algorithms are not transparent to users. They do not identify which actions are considered by these systems and do not have a clear understanding of their outputs.

To users, lists of suggestions and recommendations often seem random and meaningless. Frequently, these algorithms create categories and group items according to obscure and not mutually exclusive criteria. While these groupings may make sense from a strictly statistical or mathematical point of view, they go against the traditional intuitive way of building an information architecture.

One of the problems pointed out by Budiu [8] is the so-called black-box model. Users necessitate developing a mental model of how the system works, and it needs to be clear about how they can change the outputs. Users perceive the system as a black box with cryptic inputs and outputs out of control.

Not only are people unaware of the universe of possible inputs, but there is a significant delay between input and output, which creates higher confusion. Occasionally, imperfect metrics of relevance cause items of interest to be hidden from the user. This problem also causes the order of items on a list neither understandable nor predictable. Also, Budiu [8] says, items of low relevance are presented to the user (low precision) who will need to ignore them. However, the cost of ignoring a weak recommendation may be high depending on each specific service (in Spotify, this cost is higher than in Netflix, for example).

Repeated recommendations, as well as session-specific thumbnails, descriptions and headings, often increase the cost of interaction. Furthermore, the practice of homepage customisation, according to a session or device, restricts the learning of the layout, reducing usability.

Budiu [8] presents us with some guidelines that should be considered by the machine-learning algorithm based interface developers in order to increase user satisfaction and experience quality. These are some of them:

(i) Be transparent - Inform people about which actions are considered by the algorithm to help the user build a transparent mental model of interaction.

(ii) Provide easy controls over the output - Give the user the means to reorganize the output in a way that is more relevant or familiar to him.

(iii) Do not repeat content items within several categories - in order to decrease the cost of interaction.

Harley [10] also suggested a list of guidelines to support user experience (UX) in ML-based recommendation systems:

(i) Transparency concerning data source - The recommendation system should be specific and clear about what user data is tracked and processed to generate personalised lists. This information indicates that the content is unique, individualised to the user. A clear explanation (for example, showing that it was based on the viewing history) helps users decide if recommendations are relevant and add credibility to the system.

(ii) Highlight recommendation lists - The user always prefers suggested content in comparison to generic lists since recommendations are seen as a valuable navigation resource amid information overload. Therefore, recommendations need to be well highlighted and easy to find.

(iii) Recommendations in subcategories - One should avoid showing all recommendations allocated in an extensive “Recommendations” category. It makes more sense for users to search for content in specific subcategories, especially in environments with huge inventories, such as e-commerce or entertainment sites (Harley) [10].

(iv) Allow users to improve recommendations - Loyal and frequent users will want to interact to make recommendations more relevant. Therefore, it is necessary to provide an effective method of entering feedbacks or editing the data used to create the recommendations. An example is the possibility to edit past actions, such as deleting items from browsing history or previous purchases. This task empowers users to instruct the algorithm to forget behaviours irrelevant to their profile. Netflix users view their past activities by visiting the viewing history. In this platform, users are allowed to remove items they consider atypical from their profile.

(v) Recommendations update - According to Harley [10], if users choose to optimize their recommendations (by performing a rating, adding items to a favourites list or updating their profile), they should expect the effect to be immediate, especially when the feedback is negative.

4 Netflix and Its Recommendation Algorithms

Netflix's recommendation systems are made up of a collection of different algorithms and are an essential part of its business model in a growing market such as internet television. In this business, unlike cable TV or broadcasting, choosing what to watch and when to watch is the main challenge, as humans are known to be deficient in dealing with large volumes of information and choices, and are quickly overwhelmed.

For example, according to Gomez-Uribe and Hunt [4], a typical user loses interest within 60 to 90 seconds after reviewing around 10 to 20 titles (only 3 in detail) and browsing a maximum of two screens. The challenge is to make sure that on both of these screens, each user finds some video they consider appealing and understands why it might be interesting.

Gomez-Uribe and Hunt [4] say that personalisation, along with a recommendation, is critical to Netflix's business model and has been able to save $ 1 billion a year by avoiding subscriber cancellation and high subscriber turnover. Each user experience is customised in a variety of ways: suggested videos and their rating, video rows and pages arrangement, and even the poster thumbnails display. For in-depth customisation, Netflix combines several different algorithmic approaches to address each unique member, stated Netflix Research [12].

For Kathayat [6], traditional TV networks use standard demographic methods such as age, race, gender, or geography to plot their segmentation. Netflix tracks its members' viewing habits to define, through clustering processes, about 2,000 "taste communities" that play a central role in generating its recommendations. Algorithms create recommendations based on the assumption that similar visualisation patterns represent similar tastes.

The recommendation algorithms used in combination by Netflix are as follows:

(i) Personalised Video Ranker (PVR) - The Netflix interface is usually made with 40 lines on the homepage (depending on the access device) with up to 75 videos per line. The videos on each line usually come from a single algorithm, the PVR. This feature sorts the entire catalogue of videos (or subsets of genres) customised for each user member. This way, the same genre line displayed to different users can present completely different videos. PVR can be combined with some non-personalised popularity indicators to generate more relevant results (Gomez-Uribe and Hunt) [4].

(ii) Top-N Video Ranker - This algorithm aims to find the best-personalised recommendations from the entire catalogue for each member and show them in the Top Picks line.

(iii) Trending Now - Short-term time trends accurately predict videos that users will watch and can be combined with a certain amount of customisation. These trends are of two types: periodically recurring events (for example, a romantic video showing trend around Valentine's Day; and short-term one-off events such as a decisive soccer game covered by traditional media may increase interest in videos about sports competitions or the World Cup. According to Kathayat [6], Netflix also applies a type of near-real-time (NRT) recommendation because it knows that batch processing, used by most Internet companies, it is not agile enough to handle highly time-sensitive scenarios.

(iv) Video-Video Similarity - Lines classified as Because You Watched (BYW) are a type of item classification, grouping them according to a single video watched by individual user members, and are listed by the similarity algorithm (also called "sims"). The choice of which BYW lines enter the home page, as well as their specific subset of videos, are defined in a customised manner by the BYW algorithm (Gomez-Uribe and Hunt) [4].

(v) Continue Watching - Gomez-Uribe and Hunt [4] explain that this is the agent who places the videos in the Continue Watching line based on estimation to continue viewing or abandoning a not so exciting show. It works with data that measures the time elapsed since viewing, the point of abandonment (beginning, middle or end of the video), as well as which devices were used.

(vi) Evidence - This type of algorithm is responsible for the additional information of each program. They decide whether to show the synopsis, cast, awards, and other metadata, including which of the poster arts, among the various possible layouts, will be presented to represent the video. According to Kathayat [6], for each title, different images can be flagged, using taste communities as a starting point for choosing it. These images can be hundreds of millions and are continually tested among the subscribers base. Creating these designs employs machine learning techniques.

(vii) Page Generation: Row Selection and Ranking - This page generation algorithm employs the output of all previous algorithms to create recommendation pages taking into consideration the relevance to each member. One challenge is handling users' mood fluctuations in each session in addition to the fact that more than one family member may share an account. According to Gomez-Uribe and Hunt [4], this problem is addressed by Netflix with diversification. By producing a diverse selection of rows, the Page Generation algorithm allows the user to skip items suitable for another occasion, or another family member, and identify something relevant to her.

(viii) Search - On Netflix, over 80% of what users see comes from recommendations, said Basilico [13]. The remaining 20% comes from keyword search, an experience that relies on a set of specific algorithms and also involves recommendation issues (Gomez-Uribe and Hunt) [4]. An algorithm can try to find video titles that match a given word search, even if incomplete. Another algorithm will try to predict interest in a concept or genre of a program, and a third algorithm will try to find video title recommendations for a particular concept sought. Sometimes the user may search for specific videos, genres or actors that are not available in the catalogue, so the algorithm will recommend other programs that resemble the searched title.

Summarily, recommendation and personalisation are essential features of Netflix's business model, as it allows each member to see a different cutout of content tailored to her interests. However, Gomez-Uribe and Hunt [4] admit that there still are some issues regarding Netflix's recommendation algorithms. These include the goal of developing global algorithms capable of managing the full complexity of different languages, licenses and specific catalogues. Another challenge is account sharing, where multiple individuals with different tastes can use a single account. The algorithms must be able to provide useful suggestions to anyone in the family viewing at any given time.

According to Basilico [13], recommendation problems have not yet been solved because each user is unique and has a wide variety of interests. What people want to watch is dependent on the context and their mood, and data scientists still do not have devices capable of reading users' minds.

Experience personalisation is the next stage in the evolution of personalisation approaches. At the level of interaction, it is necessary to customize it through a more adaptive user interface, which can present more or less information depending on the user's needs, affirmed Basilico [13].

We also conjecture that there is a certain amount of distrust by users. They worry about which interaction data is tracked, with whom this data is shared, and what criteria are used to generate the recommendations due to the lack of transparency of ML algorithms.

5 Research Method: User Interviews

According to Courage and Baxter [14], one technique often used to study the user experience is the interview. An interview is a guided conversation in which one person seeks information from another. There are a variety of types of interviews, depending on restrictions and needs. A semi-structured interview is a combination of the structured and unstructured types. The interviewer begin with a set of questions to answer but can deviate from them from time to time.

The authors followed the idea that interviews are appropriate if the inquiry is looking to collect information from a small sample of the population. A screening questionnaire was used among students to select participants. Being a paying user and an experienced Netflix member were selection requirements.

We applied one-on-one semi-structured interviews to a small sample of design and communication students from three different universities in Rio de Janeiro, Brazil, with distinct characteristics and backgrounds:

(i) Universidade Federal Fluminense (UFF) is a federal university with a recently created design course within the engineering school;

(ii) Escola Superior de Desenho Industrial (ESDI) is a state university and the first design university of South America;

(iii) Faculdades Integradas Helio Alonso (Facha) is a private university with a bachelor's program in social communication.

The participators selected for the conversation are undergraduate or graduate university students aged 19 to 34. They have been paying members of Netflix for over five years, managing personal profiles.

They access the internet every day and use applications such as email, news, social networks, instant messages and entertainment, among others. They consume streaming videos using smart TVs, notebooks or cell phones, via Wi-Fi or downloaded content, at home or on the bus while shuttling to college. They are also frequent users of Netflix platform.

We asked students what influences the discovery of new video content. Some of them answered that they notice and like Netflix's recommendations. However, several respondents stated that they prefer recommendations from other sources, such as friends, social networks and specialised channels:

"I receive recommendations from Netflix and like them, but I am also very influenced by direct recommendations from friends. In social networks, many people are commenting on the series and films".

"Netflix recommendations are not usually very effective for me. I get informed on Youtube channels that talk about film production like Screen Junkies, Beyond the Trailer and Dalenogare".

"In addition to Netflix recommendations, I'm influenced by three Youtube channels: Carol Moreira, Mikann and Series Maniac".

"I watch the series that have been most reverberated or that have won awards. They're always on social networks, Twitter or Instagram."

"Close friends influence me directly. Netflix's suggestions are completely random…"

A question about wether users could identify where their custom recommendations are was addressed. In that case, they gave conflicting answers. We concluded that they were not sure where their personalised recommendations are on the home page:

"Sometimes, they put recommendations... But I don't know if I can identify them."

"Only when they put 'Because you watched such a thing.'"

"I think my specific recommendations appear here…" (points to 'Because you added The End of the F**ing World to your list').

"*The specific recommendations for me are not highlighted enough because the first ones that appear are 'The Most Watched' or 'The Most Popular.' That is what everyone is having".*

The interviewed users showed that they do not have an adequate mental model, when asked how Netflix's algorithms work. Many were unable to define accurately what actions the algorithm considers to create the lists of recommendations. The answers revealed that there is a "black box" feeling:

"Man, I have no idea... The algorithm on Facebook works great. But on Netflix, I have no idea because it's so commercial."

"It's just the search and watched programs."

"I think the likes (formerly stars), I think it sees what I save on my list."

"I think it's from what I see. When it scored. (...) Maybe the trailers I watch. I don't know very well."

"Watch, watch and stay, watch and leave... Staying a long time, opening the video description many times."

"I believe they see in my history the series I watched, and the series I put on my list, and if I give positive feedback, leaving a Like there. In my perception, it collects all this... I believe that the search can also influence."

"What I say on the cell phone influences for sure. It happened several times I talk 'I Love How To Get Away with Murder' and this series was emphasised on my Netflix home."

"I don't know if it gets stuff from outside, from another site…"

Another question we asked was whether users knew how they could interact to improve the quality of their recommendations. From their answers, we realised that several respondents would not know how to do this:

"I don't know how I could do that."

"I don't know, and I don't see any alternative."

"I usually ignore bad recommendations. (...) I spend a lot of time away from Netflix for not receiving a relevant recommendation the moment I finish watching a series."

"If I watched a show and gave negative feedback... I can't think of another way."

"Maybe by giving Dislike even without having watched it. In fact, I never got to do that." (When giving a Dislike in a video, the interviewee did not realize that the thumbnail had become black and white).

"Dislike is unfair. I may not have liked the Queer Eye series, but from the moment you give a Dislike, the algorithm will stop showing you LGBT content".

After, we asked respondents how they understood the matter of their data protection, and the likelihood of potential threats to privacy, due to the use of navigational data. In these responses, some users manifested distrust and reinforced the idea of the "black box." Some conspiratorial theories appeared, including political implications:

"I am very suspicious of these things. (...) On the cell phone, sometimes you're talking about something, and then a lot of advertising starts to appear."

"I watched 'The Edge of Democracy"[1] *and loved it. But I'm not sure if they're gonna create a list of all the people who were interested in the Netflix left-wing movie... Got it?"*

"All social networks are a way to map people in a very clear way. Not only the film ones but... I think we give too much information! But we have no choice: either we give the information or we don't live in the modern world!"

"My data is all sold! I'm conscious of that, and I've already accepted it. About two or three years ago, people started talking about this, about how Facebook was selling our data while Marc Zuckerberg was investigated. Then, I had a credit card cloned…"

"Everyone feels invaded and surveilled by the algorithms. This is the truth, in everything you access on the internet. Not on Netflix because the algorithm seems a little random to me".

"Through the algorithm they work on the social groups and the niches... I think this is dangerous or wrong because you are limiting free access to information."

---

[1] Brazilian political documentary nominated for the Oscar 2020.

"If I search for something I want to buy on the internet, then this product appears in different forms and different prices in ads everywhere. This is the feeling of invasion of privacy that I have…"

"I keep thinking and thinking: man, everything I do these guys from those big companies know. Damn, they're watching me!"

"This is the invasion of privacy. At first, you get scared, but then you see that in a way, it's cool."

"I feel a little uncomfortable: if I spend the whole day searching a shopping site, the recommendations of these products appear on Instagram, Facebook or Youtube for me soon. But in Netflix's case, it brings comfort: more content that I like."

"I would be sad if Netflix made my data available to another company."

"To tell you the truth, I never stopped to think about it."

"I don't know, one thing I don't like is that my mother finds out what series I'm watching..."

6 Conclusions and Notes for Discussion

During our user interviews, we examined the content consumption practices on the Netflix platform, identifying some aspects of their interaction with recommendation algorithms. We intended to understand more in-depth about the quality of these experiences.

This investigation tried to respond to the following questions: What mental model users created concerning ML-based interfaces? Users were able to form a satisfactory mental model about how recommendation systems work? Are users satisfied with their custom recommendations? Can users interact with the recommendation algorithms in order to improve their outputs and better balance them for their purposes?

As Budiu [8] stated, in order to increase user satisfaction and experience quality, designers should regard the requirement that machine-learning algorithm based interfaces be transparent. Interfaces should inform users about which actions are considered by the algorithm to create recommendations lists. However, concerning the transparency of the information, we realised that interviewees had not correctly understood how the system works and have not formed an adequate mental model about tracked data, as well as did not understand how it is processed to create personalised lists. As Rocca said, explainability is quite important: users must understand why a particular item was presented to them; otherwise, they may lose confidence in the system.

Interviewees' responses suggested that there may be communication failures. Therefore, UX designers should strive to make it obvious how the system tracks and processes user interaction data. It would be advisable to achieve reasonable transparency and help users build appropriate mental models as well as deconstruct the "black box" feeling mentioned by Budiu [8].

Another issue concerns data privacy. Users have revealed a suspicion concerning algorithms and what might happen to usage data, not only in the Netflix platform but also in other services that use algorithm-based recommendation. When asked about the issue of personal data privacy, respondents broadened the spectrum of reporting their experiences to other services as if everything were part of the same system. Fears of political manipulation, restriction of freedom of information, and even credit card hacking have emerged. Additionally, an interviewee reported the concern that the streaming platform would share her data with other online services. These suspicions seem to be manifestations of the "conspiracy theories" pointed out by Lovejoy and Holbrook [11].

Likewise, users did not seem to be completely aware of how they could interfere with the recommendation system and what actions they could take to generate more relevant items lists in line with their goals and mood.

Respondents were unaware of the universe of possible inputs to improve the experience. Responses seem to indicate that UX designers should provide an effective method to encourage Brazilian users to record feedbacks or edit user data in order to create better recommendations. Frequent users do not know how to interact with recommended content to help the algorithm improve its recommendation lists to better suit their profile and personal preferences. Instead of answering the question about improvement, some students criticised the recommendation system.

Human-algorithm interaction emerges as the new frontier of studies involving interaction design and interface ergonomics. This research has addressed the sensitive and emerging recommendation system issues that are often considered too new or unknown by practitioners in the field of ergonomics and user experience design. The sample of interviewed users, although limited, pointed to several aspects that impact the interaction that young Brazilians had with streaming content recommendation algorithms such as Netflix's.

Recommendation problems have not yet been solved since each user is unique and has a wide variety of interests (Basilico) [13]. Therefore, UX designers must keep in mind that they still have a lot to learn and to contribute to the reinforcement of the user's control over technology during the interaction with widely adopted machine learning algorithms.

References

1. Bengio, Y.: Yoshua Bengio on intelligent machines. https://youtu.be/ePUSElR0o9o, last accessed 2019/12/27.
2. Pandey, P.: The Remarkable world of Recommender Systems. https://towardsdatascience.com/the-remarkable-world-of-recommender-systems-bff4b9cbe6a7, last accessed 2020/01/28.
3. Anderson, C.: A Cauda Longa: Do mercado de Massa para o Mercado de Nicho. Elsevier. Rio de Janeiro (2006).
4. Gomez-Uribe, C.A.; Hunt, N.: The Netflix recommender system: Algorithms, business value, and innovation. In: ACM Transactions Manage. Inf. Syst. 6, 4, Article 13. ACM, New York (2015).
5. Rocca, B.: Introduction to recommender systems. https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada 1/22, last accessed 2019/12/19.
6. Kathayat, V.: How Netflix uses AI for content creation and recommendation. https://medium.com/swlh/how-netflix-uses-ai-for-content-creation-and-recommendation-c1919efc0af4, last accessed 2020/01/15.

7. Beel, J., Langer, S., Nürnberger, A., Genzmehr, M.: The Impact of Demographics (Age and Gender) and Other User Characteristics on Evaluating Recommender Systems. In: Aalberg, T., Dobreva, M.,Papatheodorou, C., Tsakonas, G., Farrugia, C. Proceedings of the 17th International Conference on Theory and Practice of Digital Libraries. 400–404. Springer: Valletta, Malta (2013).
8. Budiu, R.: Can Users Control and Understand a UI Driven by Machine Learning? Nielsen Norman Group, https://www.nngroup.com/articles/machine-learning-ux/, last accessed 2020/01/02.
9. Harley, A.: Individualized Recommendations: Users' Expectations & Assumptions. https://www.nngroup.com/articles/recommendation-expectations/, last accessed 2019/11/28.
10. Harley, A.: UX Guidelines for Recommended Content. https://www.nngroup.com/articles/recommendation-guidelines/, last accessed 2019/11/29.
11. Lovejoy, J., Holbrook, J.: Human-Centered Machine Learning. https://medium.com/google-design/human-centered-machine-learning-a770d10562cd, last accessed 2019/12/10.
12. Netflix Research.: Machine Learning: Learning how to entertain the world, https://research.netflix.com/research-area/machine-learning, last accessed 2019/12/28.
13. Basilico, J.: Recent Trends in Personalisation: a Netflix Perspective. ICML 2019 - Adaptative and Multi-Task Learning Workshop, https://slideslive.com/38917692/recent-trends-in-personalization-a-netflix-perspective, last accessed 2020/01/10.
14. Courage, C., Baxter, K.: Understanding Your Users: A Practical Guide to User Requirements Methods, Tools, and Techniques. Morgan Kaufmann, San Francisco (2005).