The secondary ethical layer: A solution for the ethical issues of recommender systems

Author: Stijn Valkenburg
Student number: 5868513
Supervisor: Michael Dé
Secondary supervisor: Natasha Alechina
Date: April 16th 2021
Study: Bachelor Artificial Intelligence, Utrecht University
EC: 7,5
Words: 7368

Abstract

Recommender systems are all around us; they can be found in news applications, YouTube, Netflix, the healthcare industry, and e-commerce. These recommender systems are influencing our choices and the information that is presented to us. This makes it crucial to think about the ethical consequences of these recommendations and possible solutions to ethical issues. In this thesis, we have identified the main ethical challenges of recommender systems, and we looked at one specific, promising solution called the secondary ethical layer. The secondary ethical layer is a general ethical filter which filters out any unethical recommendations based on cultural and personal preferences while also taking into account all the different stakeholders on which recommendations can have an effect (such as the user, provider, system and society). We have found that this solution can solve some ethical issues, specifically with regards to inappropriate content, unfairness (biases) and issues for society. It does not solve problems such as the lack of opacity and some privacy issues within recommender systems. This thesis identifies different key elements of the ethical layer and creates the fundaments on which a practical solution can be built.

Introduction

Ethical Layers as a solution to ethical problems of recommender systems (RS)

We constantly interact with and are influenced by recommendations systems (hereinafter referred to as RSs and also known as recommender systems). We can find them in news applications, YouTube, Netflix, advertisements, e-commerce websites (Paraschakis, 2018) and the healthcare industry (Sezgin and Özkan, 2013) and many other places (Paraschakis, 2016). RSs are algorithms that make suggestions about what a user may like, such as a specific movie or product. RSs are ubiquitous, and there is much research on developing more advanced and efficient systems. These projects are done mainly by businesses and are driven by online commerce and services, where the emphasis tends to be on commercial objectives. However, RSs have a much broader impact on their users, providers and even society. They shape our preferences and guide (sometimes critical) choices, both for individuals, groups, and society (Milano et al. 2020). The considerable influence these systems can have on our lives makes it crucial to think about the ethical problems caused by RSs, like inappropriate content, privacy, social effect, encroachment on individuals, and more.

RSs are a specific subgenre of the field of Artificial Intelligence. The suggestions they make are primarily based on machine learning techniques, but they also use many other tools. RSs are also responsible for a significant part of the infosphere of their user (Floridi, 2013). A RS influences which information a user is provided with and, more importantly, which not. This makes the RS part of different ethics branches, like Computer Ethics, AI Ethics, and information ethics. These fields are all relatively new, and there is no consensus between them on addressing ethical issues. Some still look at the field of computer ethics as if it is a practical subject. One of the reasons for this could be that computer ethics is at too much of a crossroad of technical matters, moral and legal issues, socials, and political problems to be someone's own game (Floridi, 1999). At the same time, many institutions, governments, and companies are creating guidelines for AI, information, and therefore also RSs. However, as explained by Hagendorff (2020), there is a big difference between those guidelines. He states that they use different ethical frameworks based on a virtue ethical approach, some more Kantian and some utilitarian.

If we zoom in on RSs from an ethical perspective, we can see that it has a consequentialist view of 'good' and what is not. The primary way to measure the 'goodness' of a recommendation is by using its utility. Milano et al. (2020) state the following: "at least two variables are morally relevant: actions and consequences. Of course, other things could be relevant, in particular intention. However, for our purpose, the aforementioned distinction is all we need". Both the actions and consequences are being captured in the utility of the RS. For this reason and simplicity, we will focus on a utilitarian view in this thesis. The utilitarian view already is a widely discussed topic in the ethical field and we will not try to find and present in this thesis a uniform definition that captures this view. From this utilitarian viewing points, Milano et al. (2020) have made a suggestion for 6 categories of ethical issues, namely: inappropriate content, opacity, privacy, unfairness, encroachment on individuals and society.

This thesis aims to look at how we could define and structure the ethical problems of RSs, and the effect a secondary ethical layer in RSs has on solving or minimising the effects of these problems. In order to do this, we should first define what a RS is and how the different kinds of RSs work. Afterwards, we will describe the ethical frameworks in which RSs work and how they compare against the other strategies in information and AI ethics. Thirdly, we will explain what a secondary ethical layer is and why it could help solving some ethical problems with RSs. After defining this, we will look at the different ethical problems within the framework proposed by Milano et al. (2020) and how, if possible, a secondary ethical layer can solve these problems.

Chapter 1 – Recommender systems

What are RSs?

RSs can, in essence, be seen as a form of information filtering systems (Pfaff, 2021). Alternatively, as stated by Rodriquez and Watkins (2009), a RS is an information filtering tool that matches individuals to resources of potential interest. More formally, RSs are systems that take data/information from a user as input and a prediction of the users' rating of an item as output. These systems will then predict how the user would rank the set of items. As suggested by Milano et al. (2020), three parameters are essential to make a RS operational. These parameters are:
a) the space of possible options/items,
b) the definition of what a good recommendation would be (the utility of the recommendation)
c) how to evaluate the RS's performance.
What the values of these parameters are, is very dependent on the Level of Abstraction.

RSs are used in a wide variety of different applications. The first RS appears in 1990 from Karlgren and was called ‘Algebra for Recommendations’ (Karlgren, 1990). This kind of system was based mainly on some matrix calculations, whereas nowadays, we mostly use big data-driven systems to train machine learning (like Artificial Neural Networks, also known as ANN's) to calculate the expected likeness for the user. However, behind these methods are the same kind of techniques, and most often the systems are based on content-based systems (CB) or/and collaborative filtering (CF) methods or a combination of these two methods. These systems are called hybrids (Burke, 2002; Milano et al., 2020).

A collaborative filtering (CF) system can be seen as making a recommendation based on collaborative information from multiple similar users. The RS compares users based on their earlier decisions or ratings. These systems assume that users who made similar choices will continue to do so in the future. So, it will recommend an item that is liked by those who are like the user. CF systems are often found in e-commerce or social media platforms. In contrast stand contest-based (CB) systems. Where the CF system looks for similarities between users, the CB systems are looking for similarities between items. For example, when items are described with keywords, the RS is looking for items sharing the same or related words. A higher amount of shared keywords hints at a closer relation. If a user then rates one product highly, it will recommend a similar product. The combination of them is called hybrid RSs. All three systems have their advantages and disadvantages. More can be found in the papers of Pfaff (2021) and Burke (2002). There are also three other techniques: demographic, utility-based and knowledge-based systems, but these are more rarely used than the former two, so we will not discuss them (Burke, 2002).

In theory, we should not be limited by the kind of RS we are using for the solution we are suggesting. However, to this date, we do not have enough knowledge about ethical problems arising in other RSs. Therefore, some ethical problems discussed in this thesis might not be relevant for other techniques or other issues will solely occur when other RSs are used (Pfaff, 2021).

Level of Abstraction

In the previous section we have defined how a RS works and what techniques are used. Now it is essential to talk about the Level of Abstraction (hereinafter referred to as LoA). The LoA is a set of observables. An observable is an interpreted typed variable, that is a typed variable together with a statement of what feature of the system under consideration it represents (Floridi, 2008). Let us take you and your friend tasting wines as example. You may have different ways of measuring wine (nose, robe, colour etc.), but the level you talk on is still the same. You are talking about the quality of the wine from a tasting, personal perspective. However, if an auction house is trying to sell a bottle of wine from 1930 for the highest price, they do not use variables like nose, robe, or colour to talk about the quality of the wine. In this case, you are at a different LoA (Floridi, 2008). The same holds with RSs. For example, when we are looking at a RS for e-commerce, we have a space of items (A) consisting of items that could be bought. A good recommendation (B) could be to suggest an item that eventually is being bought. A way to evaluate the recommendations (C) could be the click-through rate or just a binary value if something is being bought or not. When looking at a different example, say YouTube, we see a different LoA. Here the space of items (A) could be seen as all the possible videos on YouTube, a good recommendation (B) could be a suggestion of a video which the user then is going to watch, and the evaluation function (C) could be the time the users stays on the platform (Milano et al. 2020).

It is crucial to understand the LoA when working with different RSs (and all information systems). Some ethical problems only occur in specific LoA's. There is a big difference between RSs recognising health issues when being used by a doctor who is figuring out what operation to do and a health RS for an internet user who just wants to know if the illness is serious enough to go to the hospital. Most of the literature on RSs focusses on a LoA where the internet user plays the biggest role, but there are many other possible LoA's. In this paper, we will try to generalise some ethical problems, but we will indicate whenever some problems only occur for some level of abstraction and when research is focussing on a specific LoA.

Ethical problems

There are many kinds of ethical problems with RSs and with AI systems in general. Think about seeing disturbing recommendations, biases, information bubbles or echo chambers, privacy issues and loss of autonomy. In the remaining part of this thesis, we will look at these problems within the classification from Milano et al. (2020), and we will look at a particular solution, called the secondary ethical layer. As explained earlier, we are looking at actions and consequences. Consequences and actions from ethical problems in RSs could have two possible effects on the stakeholder in a RS.

	Immediate harm	**Exposure to risk**
Utility	Inappropriate content	Opacity
Rights	Unfair recommendations, Encroachment on individuals	Privacy, Social effects

Table 1: Summary of identified issues with RSs by Milano et al. (2020)

It could lower the utility of the item thereby making the recommended items less desirable. Moreover, it could violate the rights of the stakeholder. For example, it could discriminate against the stakeholder or violate their right to privacy. We can also see a difference in the timing of the consequence. The unethical effect could happen immediately. For example, the user sees something he rather would not have seen or it could expose a risk for an unethical effect in the future. When systems gather too much data, it will not have an immediate effect but it could become a problem once the data is being compromised. Milano et al. (2020) also identified, by a multidisciplinary and comparative meta-analysis, six main areas of ethical concerns. They do overlap, but they give us some handles to look for solutions. These are presented in table 1. We will use these handles in the next chapters.

Chapter 2: Possible solutions

Now that we know more about RSs and have organised ethical challenges, we can proceed by looking at possible solutions for the problems. When finding a solution, there are basically two options. The first option is to change the RS itself. These solutions are called internal solutions. The algorithm of the RSs should be changed or should learn to be more ethical. The other option is to use external solutions. This are solutions that do not interfere with the RSs themselves but filter out any non-ethical data of recommendations. This can happen before or after the recommendation is being made by the original RS. Not all categories of ethical problems can be solved by both internal and external solutions. As we will see, opacity and most privacy problems are only possible to solve with an internal solution.

However, there are two problems with internal solutions. The first one is that businesses often do not share any information about the RS and are unwilling to change them. For them, the primary stakeholder is, most of the time, the company or platform. They, for example, profit if the user stays on the platform for the most prolonged time or buys the most products. The second problem is that for an ethical RS to work, it should include a lot of personalised features. Ethical frameworks are strongly dependent on demographic, personal and cultural preferences (Souali et al. 2011). It can be quite a burden for (smaller) companies to manage all those preferences and include them into the recommendations systems. Therefore, we will solely focus on external solutions.

External – The secondary ethical layer

One of the external solutions is suggested by Ya Tang and Winoto (2016). They suggest creating an additional ethics layer above the recommendation. This has a few advantages: it could better manage personal and cultural preferences and it does not interfere with the working of the RS so implementation can happen easily and more extensive. They call this the *two-layer ethical RS*, which they describe as “The first layer can match a target user's preferences against an item database and other users with similar interests and making suggestions *(this is the same function a standard RS has)*; the second layer is an ethical filter picking up ethically appropriate items based on a given set of ethical rules and content analysis of candidate items.” This is the basic form of the ethical layer which we will expand further.

The ethical layer could work within two places in the architecture of the RS. It could be a filter implemented by the companies to manage the data sources and filter out the items before they go into the RSs, or it could filter the results coming from the RS and only show the ethical ones. Implementing the filter before the recommendations will not solve all the problems, such as biases within the RS, and it is questionable why companies want to do this and do not want an internal solution. Implementing the filter after the recommendations is intuitively the best option. However, a concern that is being raised is that such a secondary layer will harm the RS's success and function. Some say that, for example, Netflix should give you the best possible movie and when filtering out some results, you do not get the best recommendation anymore. Although this looks true when looking at it from the end user's perspective, we can see that some suggest that the best recommendation is not the best recommendation for the end user, but for all the stakeholders (Milano et al. 2021).

This brings us to the point of the multi-stakeholder recommendation. Traditionally a RS is looking for the best recommendation for the end-user of the system (Milano et al. 2021). It searches for the right items to maximise the utility of the user. A good recommendation is a recommendation where the user is most satisfied. In contrast to this traditional user-centred approach, which is too impoverished to account adequately for the social impacts of recommendation, a new research paradigm is emerging that explicitly models multiple stakeholders in the systems. Milano et al. (2021) are identifying four different stakeholders of each RS. *Users* are the parties to whom the recommendation is targeted. *Providers* are the parties who make the options available. These are affected by the recommendations that the system makes to users, in so far as their "items" can receive more or less attention depending on how they are recommended. The *system* captures the interests of the platform on which the recommendations are generated. The *society* is systemically affected by the recommendations made by a system, for example by altering or reinforcing existing social norms.

A lot of ethical issues that are being imposed by RSs are caused by an imbalance between the different stakeholders of the RSs. Therefore, we propose that our ethical layer should have a form of a multi-stakeholder RS, where it weighs the different stakeholders against each other and is transparent about the effect to different stakeholders of each recommendation. Implementing a multi-stakeholder approach within the ethical layer is not too complicated in the most basic form. Instead of just calculating the utility for the end user, you also take in account the utility for the other stakeholders. However, one of the main issues with multi-stakeholder RSs is that we cannot easily weigh all the different stakeholders in our classical RSs. The different stakeholders are working in a different LoA (Milano et al., 2020) and identifying all the relevant stakeholders in a recommendation is often unlikely, if not utterly unfeasible. In the most straightforward cases, there may be consequences of a recommendation that affect parties in ways that are difficult to anticipate. Kermany et al. (2020) show that with two stakeholders, the providers, and the user, they effectively provide more long-tail items (unpopular items) and better fairness for the provider with a small loss of accuracy.

In the version of Ya Tang and Winoto (2016) we can identify parts of the secondary layer where we see some limitations. The first problem being that the filter only works with pre-labelled data such as movies. It analyses the content of the recommendations based on this data, which it collects from sources like the Internet Movie Database or the Motion Picture Rating. Therefore, their ethical filter only works on specific domains such as books and movies. This issue is also being addressed by Rodriquez and Watkins (2009). They suggest the usage of a web of data, where different kinds of databases are being linked. In this way, the system has a more extensive knowledge base for understanding the world and the individuals' place within it. It also extends the range of domains the ethical layer could work on. However, this has some problems in itself, with the most prominent one being that it is tough to collect these data, as it comes from different sources that differ and are not optimised for data collection. Another way at looking at the content analysing part is by implementing a new content analyser tool that should in some way be able to analyse the recommendations being made. There are many tools for analysing the multimedia content. For more details on state-of-the-art techniques, we recommend Deldjoo et al. (2020) or Karlsson and Sjovaag (2016).

Another question is how the user should be able to set up their preferences. In the version of Ya Tang and Winoto (2016) the filtering is being done based on rules. Setting up these rules are very labour intensive for the user. Therefore, Ya Tang and Winoto (2016) suggest that this should happen gradually through multiple interactions with a conversational RS. Although this would increase the user experiences, when working with more that a few rules and more complex items than movies, this will not be enough. One suggestion could be to make some pre-sets based on demographics and a choice from the user (Souali, El Afia, & Faizi, 2011). This could lower the time of the initial set up significantly and we could then specify the systems more gradually through multiple interactions. Another important aspect of our ethical layer is, that our systems should be explainable and transparent (Floridi et al. 2018).

Chapter 3: Specific ethical problems for RS and their EL solutions

Now we are more familiar with what an ethical layer is, we can ask ourselves: how could an ethical layer solve some of the ethical problems caused by RSs. As seen above, we have discussed the framework from Milano et al. (2020) as a methodology for these ethical problems, which can be found in table 1. In the following part, we will discuss each of the problem categories and figure out if an ethical layer could help solving these problems and what capacity and which parts of the ethical layer should be altered to do this.

Inappropriate content

Imagine a child using a movie RS. It will be wrong if the child sees violent or sexual content. This is one of the most tackled problems both done by business as in research. Apart from the apparent need for parental control, one might think of morally troubling examples of recommending meat-based products to a vegetarian, alcohol drinks to a religious Muslim, or tobacco products to a person who struggles to quit smoking. A problem stated in some literature is being caused by families using the same systems. We also noticed that nearly all RSs have accounts/profiles for different users, so we do not think that account sharing is a problem. (Ya Tang and Winoto, 2016)

There are different ways to solve the problem of inappropriate content. The first one would be the ethical filter proposed by Ya Tang and Winoto (2016). As seen before, they use a secondary layer based on a set of rules to form an ethical layer above the RS. As we have discussed earlier, there are some disadvantages to this approach. Another system is the eudemonic system of Rodriguez and Watkins (2009), who are suggesting a eudemonic RS whose purpose is to "produce societies in which the individuals experience satisfaction through a deep engagement in the world" (Rodriquez & Wakins, 2009)The authors think this would be achievable by creating an ethical filter based on a large, interlinked data structure. Another way to tackle this problem is by check the appropriateness of candidate items by mapping potentially harmful elements in the content (drug use, nudity, etc.) to a user's persona (gender, age, religion etc.) as Ya Tang and Winoto (2016) also explain.

We could also use demographic or geographical data to filter cultural norms to which the recommendation should hold up (Souali et al., 2011). This however does not work with the view from Paraschakis (2018) where he says that users should have full control over the filtering process. And the solution of Souali et al. (2011) does also raise some other problems of which the most prominent are autonomy, content censorship and the multi-stakeholder problem because a filter based on rules, as suggested above, will work only for the user (Paraschakis, 2018). Paraschakis (2018) says that users should have control over the filtering process.

Privacy

Users' privacy is one of the main challenges of RSs. As we have seen above, most successful RSs are based, at least partially through a hybrid RS, on collaborative filtering techniques, which depends on collecting, storing, handling, and comparing user data. Privacy risks can occur in at least four stages. Firstly, they can arise when data is collected or shared without the consent of the user. Secondly, when the data is stored, there is a risk of data being leaked, hacked, or that the data becomes subject to de-anonymisation attempts. Thirdly, there is a risk from inferences that a system can draw from the data, where users may be unaware of. For example, a user could get a recommendation that shows some sensitive information about the user. Finally, it can create a model of a user, without information of the user itself but with the usage of data from comparable users. (Milano et al. 2020; Friedman et al., 2015) There are currently three kinds of solutions being suggested: architecture (for example: storing data decentralised to minimise the change on a data leak), algorithmic (using encryption en anonymisation) and policies (for example, the General Data Protection Regulation of the EU, also known as the GDPR).

Not much work has been done on privacy and ethical layers. The solution should not prevent the RS from working, so data cannot be withheld from the RSs. An ethical layer can also obviously not interfere with how the architecture of the RS is working or what policies are being implemented, rather than being a tool that makes people, governments, and institutions aware of problems. The ethical layer could anonymise the data before it goes to the RS, but again this should not harm the recommendation. The ethical layer could also notify the user if it thinks data is being used or collected without the consent of the user. It could use the systems' opacity to look at the recommendations based on data not willingly provided by the end-user. Paraschakis (2018) suggests that the RSs should have privacy controls that the user can configure. Privacy issues also come up when making the ethical system. It should store and use data from the user to create the ethical layer, it has access to some sensitive information about their ethical preferences, so it should be highly encrypted and decentralised to minimise all privacy risk.

Opacity

The problem of opacity within the RS is not a problem which can be solved by an external solution but should be solved by an internal solution. In theory, explaining how personalised recommendations are generated for a user could help mitigate the risk of encroaching on their autonomy (Milano et al. 2020) but could also reduce the effects of all other ethical issues, by making is easier for the stakeholder to understand the reasons for making a recommendation.

However, we will not expand on other solutions for opacity because we are focusing on external solutions. There is a lot of research being done on this field. For example, Tintarev and Masthoff (2012), Germano et al. (2019) and Floridi et al. (2018).

Fairness

It is well known that RSs do have biases or unfairness in their recommendation. Fernadi et al. (2018) state that there are two primary sources of unfairness in RSs. The first one is observation bias, which results from a feedback loop generated by the system's recommendations to a specific group of users. The second one is the population imbalance, where the data available to the systems reflect existing social patterns that are expressing biases toward some groups our items. Andollahpouri et al. (2019) also talk about the concept of popularity bias, which is related in definition to the observation bias but is about the users in general. The popularity bias could be harmful for several reasons, such as that long-tail items are important for generating a fuller understanding of users' preferences. Another reason is that systems that use active learning to explore each user's profile will typically need to present more long-tail items because these are the ones that the user is less likely to have related, and where user's preferences are more likely to diverse. In addition, long-tail recommendations can also be understood as a social good. A market that suffers from popularity bias will lack opportunities to discover more obscure products and will be, by definition, dominated by a few large brands or well-known artists (Friedler et al. 2016).

As a solution, multi-side concept for fairness is being proposed. This closely relates to the multi-stakeholder RS but is slightly different in that it uses only three sides: the user and the provider and the combination. Using this taxonomy, a developer of a RS could identify how the competing interest of the different parties is affected by the recommendation. As seen before, we suggest that the secondary ethical layer implements a multi-stakeholder RS to calculate for every recommended item the effects on the different stakeholders (Burke, 2017; Milano et al. 2021).

A second way is to identify biases in the different recommendations by using a probability matrix. Both Marklin et al. (2007) and Yao and Huang (2017) suggest checking the different groups of recommendations with the different real-life groups. In this way, we could find out if some groups are recommended more often. This technique is very labour and technology intensive because it should gather information about real-life groups and probabilities. Whenever a bias is identified, an ethical layer could choose to show that a bias in on the recommendations, or it could choose to filter out biased recommendations completely.

Encroachment on individuals

A fourth ethical issue in RSs is the encroachment on individual users' autonomy, by providing recommendations that nudge users in a particular direction, by addicting them to some types of content or by limiting the range of different options (or opinions) to which they are exposed. These problems are known in different ways as filter bubbles, echo chambers or information cocoons. This could be harmful not only for the end-user, by steering their ideas and behaviours, but also could have a more significant effect on democracy (Borgesius et al. 2016). It is important to distinguish different kinds of filter bubbles: self-selected personalisation, where people actively choose which content they see, and pre-selected personalisation, where algorithms personalise the content without any deliberate user choice. However, the effects of personalised and selective news menus are different for everyone, and many people are not affected (Valkenburg & Peter, 2013). This means that the problem and solutions of encroachment are different for all the users.

Our experience of personal identity is mediated by the categories in which we are assigned in RSs. When assigned to a specific persona, we are more likely to be trapped in a bubble. Solutions to this problem include the explore and exploit paradigm (e.g., learning the world), diversity, novelty, and serendipity (De Vries, 2010). There are different exploration techniques that a RS can use, and depending on the technique, a bubble could be smaller and larger. Exploration does reduce the short-term success of the recommendations and could thereby harm the short-term revenue. However, Baeza-Yates (2020) does suggest that if more exploration is performed the tension between user experience and monetisation will diminish and that that will be good for the RSs and for a more fair and healthy digital market for users and providers.

The secondary ethical layer could handle some of these problems. If in a way it could recognise filtering bubbles and echo chambers, it could serve recommendations that are not in the bubble, give insights into the bubble, given the option to leave the bubble or create awareness of the fact that some of these items are just served because they are in a specific category. The final solution has some problem to it. One of these is that computer-generated categories do not always match with the human interpretable categories we would make. We know cat lovers and can judge if we are part of this, but we cannot understand more complex structures like "clicked two cat and specific music videos" or even categories based on mathematical vectors (Milano et al. 2020).

Social effects

The impact on RSs is one of the most complex and less discussed ethical effect of RSs. As we have seen in the previous parts, the social effects are found from every aspect of ethical problems, one of these being that news RSs and social media filter are insulating the user from exposure to different viewpoints and thereby creating a self-reinforcing bias or filter bubble that is damaging the normal functioning of public debate, group deliberation and democracy. News RSs are constantly comparing expected relevance to earlier news and diversity of news to the items. Milano et al. (2020) suggest that these tools should favour democratic norms.

Other problems posed are that RSs are fragmenting internet users, reducing shared experiences and narrowing media consumption. However, surprisingly, this does not appear in all empirical studies. Hosanagar et al. (2014) say that "Personalisation appears to be a tool that helps users widen their interests, which in turn creates commonality with others. This increase in commonality occurs for two reasons, which we term volume and product-mix effects. The volume effect is that consumers simply consume more after personalised recommendations, increasing the chance of having more items in common." On the one hand, it is questionable if this study is also expandable to other domains than music, where effects could be more significant, such as news, books and research RSs. On the other hand, we may not yet understand the effects of online fragmentation as we did not understand the effects shown in The Big Short, how over 30 years, the Americans have sorted themselves into like-minded neighbourhoods. This could happen, with the help of RSs, also online (Bishop, 2008). Another problem is being the exploitation of RSs. This is altering your item in such a way that the systems recommends it more often, but not because the content is a better fit, but because you 'hacked' the system. In this way, people can misuse the RSs and influence people through the RSs. This issue is challenging to address from an external solution, as this is an issue of the RS itself (Gielen, 2016). Other social issues are less often being addressed in the literature, such as the effect of addicting content on the efficiency of working people, the effects of recommending harmful content to people who are suicidal, and questions like who is responsible for something going wrong.

The last social issue sometimes overlooked when talking about RS is about responsibility. Allen et al. (2006) say that "the modular design of systems can mean that no single person or group can fully grasp the manner in which the system will interact or respond to a complex flow of new inputs." The gap between the designer's control, user control and the algorithm's behaviour create an *accountability gap*. This accountability gap is also difficult to solve with an ethical layer, although we could imagine that, to some extent, the external layer converts the recommendation to be in responsibility of the end-user when the end-user has the possibility to alter all the settings and preferences (Cardona, 2008; Mittelstadt et al., 2016).

The typical solutions to RSs social, ethical problems are primarily being split into two strategies: bottom-up and top-down. They prioritise either the user's preferences (and their autonomy in deciding how to configure the personalised recommendations) or they prioritise the social preferences for a balanced public arena. For the ethical layer approach, we should focus on bottom-up solutions. One of the solutions suggested is creating a persona which the user can select to view the world from a different stance. The persona can be implemented in an ethical layer, but its usability for the end-user is questionable. Interests and profiles can differ highly, as there is an endless list of interests a user can have. A second solution is creating more serendipity. Exposure could create a better debate but maybe lower the utility of our RSs. The ethical layer, working for the user, could include a feature where the user could select the amount of exploration vs exploitation he would like to experience. Furthermore, as discussed earlier, one of the most important solutions is implementing multi-stakeholder RSs. When correctly implemented, the ethical layer could include the effects on society in every recommendation it could make.

Limitations to the secondary ethical layer

We have seen that the ethical layer could become a user-centred solution to some of the ethical issues. However, there are some limitations with the ethical layer that we have not yet addressed.

A problem with the ethical layer approach could be how to connect to right LoA to each system. As we have seen above, the LoA is fundamental in RSs. It mainly creates a problem for the ethical layer when trying to create a more generalised ethical layer. It could be challenging for an Ethical layer to 'know' what LoA we are currently working with. On the other hand, this question is equally hard for an ethical layer as for the RS itself. This problem grows when we are looking at RSs for health care of legal industries, where we can see completely different ethical problems occurring (Floridi, 2008). A solution to this could work around classifying different LoA to different RS’s, both manually but also with the help of machine learning.

A second problem may be the input of the user. First, we need to assume that the end-user is capable of selecting his preferences (some therefore suggest that parents should be able to alter their children's settings). Secondly, we constantly find that the end-user should be able to alter all possible settings of the systems and should have insight into the system's complete functioning. To do this, the user should gather knowledge in the working of the system. After that, the user should spend time setting up the preferences and alter the setting when the users' ethical preferences change. Ya Tang and Winoto (2016) suggest that this should go gradually through a conversational RS.

Discussion

As we have seen, we do think positively about the possibilities of a secondary ethical layer. A discussion that keeps popping up is about the goal of a RS. From the perspective we take, we need to view the RS as a tool that influences us all, and therefore we should study the RS as the cause of some ethical problems. However, some say that the RSs are part of services that companies provide and argue that we should look from a broader perspective to these ethical problems. They say that we cannot view RSs apart from the service and platform they exist in, and therefore we should not approach the ethical issues and solutions to the issues from the RS stands but rather to the service as a whole. Although we think companies should work on these ethical problems in the complete scope of their services and products, we think that some of the ethical issues are raised not by services alone, but by the combination of all different RS working 'together'. When one product is changed, it does not solve the problems. Therefore, we want to explore broader approaches and look at the things that are similar between different services. We suggest future research to look into the complex environment of the products and combine information and knowledge about design, algorithms, and content to create a broad view of what influences these ethical issues.

Further research should also specifically focus on creating more empirical data on RSs. We find many theories about RS's effects, but there is a significant lack of data on some topics. The first topic where more data should be valuable is how the user experiences the RS and how the users want some issues to be solved. As we saw with filtering bubbles, some studies say that they do not influence the user as much as some think. Another part where more empirical studies should take place, is the part of society. The effects on society are difficult to measure, as some effects do not occur immediately and can take years to develop or are not that visible. Nonetheless, these effects have an enormous impact on our lives. They can shape the directions we, as a society, are moving to and therefore should be researched thoroughly.

Another discussion point could be the technical challenges of creating a multi-stakeholder RS. To calculate a RS's effects on all four stakeholders could require an enormous amount of knowledge about the world, humans, social structures, and domain-specific structures. As far as we know, there are no easy ways to implement software tools available to manage this knowledge. However, as the technology develops, more knowledge of AI will become available, and we will capture more and more of the consequences. In the meantime, we can also focus on specific domains to begin with, where it would be easier to explain the consequences (Milano et al., 2021).

The last note to make is that most of the ethical problems in this paper are problems for an end-user, a person in society. We do not look at ethical problems of RSs of AI from the perspective of institutions or governments. We do not talk about ways to improve policies that may make suggestions on RSs less relevant. Governments do have the capacity to create policies that make some of the problems less relevant or even solve these issues (Floridi et al., 2018).

Conclusion

In this thesis, we looked at several ethical issues with regard to RSs. These issues could be categorised into six groups: social effect, privacy, opacity, encroachment on the individual, unfairness, and inappropriate content. As we have seen, these problems vary in their solutions. In this paper, we have explicitly looked at possible external solutions for these issues (and not internal solutions). We found a few key elements that an external solution should have. First, it should be a multi-stakeholder system, where it can weigh off between the different stakeholders where the effects of the RSs can occur. Secondly, it should be possible for the end-user to modify all possible settings and preferences of the systems to match personal and cultural preferences and keep the user in control of their life choices and responsibilities, influencing the user's motivation happiness. Thirdly, we should create a system that is transparent, explainable, and very privacy focused.

An external ethical layer as this could solve some, but not all, of the problems of the RSs. It can filter out any inappropriate content by creating a rule-based or a data-based filtering layer that does not recommend inappropriate content. It could also help with the fairness of the RSs. It can recognise biases in recommendations and then either filter them out or notify the user of the biases. A secondary ethical layer could also solve some issues regarding the encroachment on individuals by recognising information bubbles or echo chambers and notify the user of this, and it could promote long-tail items more by filtering out some popular items. However, we found that explaining bubbles and echo chambers is more challenging because it could be challenging for a user to understand the categories or bubbles. It is also questionable to what extent encroachment is a problem, as some users are not affected by encroachment and increasing the exploration would lower the effectiveness of the recommendation. Another problem where an ethical layer could be of help is social effects. An ethical layer could filter some recommendations based on its effects on society through multi-stakeholder calculations. It could also have persona's where it can use settings and filtering from different persons in society and thereby extend the user's view. Although these are good solutions, it does not grasp the complete scope of RS's social-ethical issues. Two categories of problems cannot be solved using external solutions (and thereby the secondary ethical layer). These are most privacy problems and the problem of opacity.

To summarize, the secondary layer would be a valuable tool to explore. It could help identify some issues and create awareness within users. It could be applied generally, which is a great advantage. However, some work needs to be done to gather more information on developing parts like the multi-stakeholder calculations and the tools to analyse content.

Bibliography

Andollahpouri, H., Mansoury, M., Burke, R., & Mobasher, B. (2019). The Unfairness of Popularity Bias in Recommendation. *13th ACM Conference on Recommender Systems.* Copenhagen.

Baeza-Yates, R. (2020). RecSys 2020. *Virtual Events*.

Bishop, B. (2008). *The big short.* New York: Houghton Mifflin.

Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., de Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Journal on internet regulation*.

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction*, 331-370.

Burke, R. (2017). Multisided Fariness for Recommendation. *Workshop on Fairness, Accountability, and Transparency in Machine Learning*.

Cardona, B. (2008). Healthy ageing policies and anti-ageing ideologies and practices: On the exercise of responsibility. *Medicine, Health Care and Philosophy*, 807-816.

De Vries, K. (2010). Identity, profiling algorithms and a world of ambient intelligence. *Ethics and information technology 12*, 71-85.

Deldjoo, Y., Schedl, M., Cremonesi, P., & Pasi, G. (2020). Recommender Systems Leveraging Multimedia Content. *ACM Computing Surveys Vol 53, No. 5*.

Floridi, L. (2005). Information Ethics, its nature and scope. *ACM SIGCAS Computers and Society*.

Floridi, L. (2008). The method of levels of abstaction. *Minds and Machines*, 303-329.

Floridi, L. (2013). *The ethics of information.* Oxford: Oxford university press.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines* , 689-707.

Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. (2016). On the (im)possibility of fairness.

Friedman, A., Knijnenburg, B., Vanhecke, K., Martens, L., & Berkovsky, S. (2015). Privacy Aspects of Recommender systems. In F. Ricci, R. L, & B. Shapira, *Recommender systems Handbook* (pp. 649-688). New York: Spring science + Business Media .

Germano, F., Vincenc, G., & Le Mens, G. (2019). The few-get-richer: a surprising consequence of popularity-based rankings. *proceedings of The Web Conference (WWW 2019).*

Gielen, M. (2016, April 14). *7 Expert Tips: How to Get YouTube to.* Retrieved April 5, 2021, from Tubularinsights.com: https://tubularlabs.com/blog/7-expert-tips-youtube-suggested-videos/

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines*, 99-120.

Hosanagar, K., Fleder, D., & Lee, D. (2014). Will the Global Village Fracture Into Tribes? *Management Science 60* , 805-823.

Karlgren, J. (1990). An Algebra for Recommendations. *The systems development and artificial intelligence laboratory working paper No 179*.

Karlsson, M., & Sjovaag, H. (2016). Content Analysis and Online News. *Digital Jounalism*, 177-192.

Karpati, D., Najjar, A., & Ambrossio, D. A. (2020). Ethics of Food Recommender Applications. *AIES*.

Kermany, N. R., Zhao, W., Yang, J., Wu, J., & Pizzato, L. (2020). An Ethical Mutli-stakeholder Recommender System Based on Evolutionary Multi-Objective Optimization. *IEEE Internation Conference on Services Computing (SCC)*.

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

Milano, S., Taddeo, M., & Floridi, L. (2021). Ethical aspects of multi-stakeholder. *The information society*, 35-45.

Mittelstadt, B. D., Allo, P. T., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*.

Paraschakis, D. (2016). Recommender Systems from an Industrial and Ethical. *10th ACM Conference on Recommender Systems*, 463-466.

Paraschakis, D. (2018). Algorithmic and ethical aspects of recommender systems in e-commerce. Malmö: Studies in Computer Science No 4.

Pfaff, H. (2021). *A Survey on Current Recommender Systems.* Frankfurt: Frankfurt University of Applied Sciences.

Rodriquez, M., & Wakins, J. H. (2009). Faith in the Algorithm, Part 2: Computational Eudaemonics. *KES 2009: Knowledge-Based and Intelligent Information and Engineering Systems* , 813-820.

Souali, K., El Afia, A., & Faizi, R. (2011). An automatic ethical-based recommender system for e-commerce. I*nternational Conference on Multimedia Computing and System*, 1-4.

Su, X. K. (2009). A survey of collaborative filtering techniques. *Advances in artificial intelligence*.

Tintarev, N., & Masthoff, J. (2012). Evaluating the effectiveness of explanations for. *User Modeling and User Adapated Interaction 22*.

Valkenburg, P., & Peter, J. (2013). The differential Susceptibility to media effects model. *Journal of Communication 63*, 221-243.

Ya Tang, T., & Winoto, P. (2016). I should not recommend it to you even if you will like it: the ethics of recommender systems. *New Review of hypermedia and multimedia*, 111-138.