KINGA SORBÁN

Ethical and legal implications of using AI-powered recommendation systems in streaming services

Recommendation engines are commonly used in the entertainment industry to keep users glued in front of their screens. These engines are becoming increasingly sophisticated as machine learning tools are being built into ever-more complex AI-driven systems that enable providers to effectively map user preferences. The utilization of AI-powered tools, however, has serious ethical and legal implications. Some of the emerging issues are already being addressed by ethical codes, developed by international organizations and supranational bodies. The present study aimed to address the key challenges posed by AI-powered content recommendation engines. Consequently, this paper introduces the relevant rules present in the existing ethical guidelines and elaborates on how they are to be applied within the streaming industry. The paper strives to adopt a critical standpoint towards the provisions of the ethical guidelines in place, arguing that adopting a one-size-fits all approach is not effective due to the specificities of the content distribution industry.

Keywords: *recommendation system, streaming service, audiovisual media, fundamental rights*

Author Information

Kinga Sorbán, University of Public Service, Institute of the Information Society
https://orcid.org/0000-0002-9288-7897

How to cite this article:

Sorbán, Kinga. "Ethical and legal implications of using AI-powered recommendation systems in streaming services."
Információs Társadalom XXI, no. 2 (2021): 63–82.
https://dx.doi.org/10.22503/inftars.XXI.2021.2.5

1. Introduction

In 2007, a new kind of service was introduced in the United States that allowed members to watch movies and television shows instantly in an online environment (URL 1). The service was called Netflix and initially launched with 1000 titles and with the aim to become a competitor of traditional DVD-rental services. As we now know, the new business model revolutionized the entertainment industry, allowing users to access a massive catalogue of audiovisual works from the comfort of their couches. Demand has grown massively year on year, with Netflix re 203.66 million subscribers in 2021 (URL 2). The service has not only proven to be a wildly successful business model, but streaming content online has become a cultural phenomenon that has even given rise to its own slang terminology, consisting of terms such as 'Netflix and chill' and 'binge-watching' (URL 3). Binge-watching is a term that refers to the phenomenon of watching multiple episodes of a television programme in rapid succession (URL 4). This is a common practice; according to a survey conducted in 2013, 61% of Netflix users binge-watch TV series regularly (URL 5). Several research studies have indicated that binge-watching can be a harmful phenomenon at the level of the individual, stating that such a viewing pattern may lead to addiction symptoms (Riddle et al. 2018) similar to "other behavioural addictions, such as loss of control and pleasure anticipation" (Forte et al. 2021, 1) but also depression and polarization. However, the interest of the streaming service providers is to fuel binge-watching to enhance users' screen time and to develop stronger user engagement, all with the purpose of realizing more revenue. Among various other tools, stronger user engagement is achieved by recommending more content to watch; preferably content that spikes the individual user's interest, leading to their further consumption. It is not a coincidence that when we watch content on a platform (not only on Netflix but on Hulu, Amazon Prime and even YouTube) we keep bumping into other interesting content. Sometimes we may feel that service providers are reading our minds and know exactly what we want (or what we think we want). However providers are not using some mind-reading magic, they are using something perhaps even better: recommendation engines. Recommendation engines are commonly used in the entertainment industry to keep users glued in front of their screens. These engines are becoming increasingly sophisticated as machine learning tools are being built into ever-more complex AI-driven systems that enable providers to effectively map user preferences. The utilization of AI-powered tools, however, has serious ethical and legal implications, and not just exclusively limited to the field of content distribution. Some of the emerging issues are already being addressed by ethical codes, developed by international organization and supranational bodies. The present study aimed to address the key challenges posed by AI-powered content recommendation engines. Consequently, this paper introduces the relevant rules present in the existing ethical guidelines and elaborates on how they are to be applied within the streaming industry. The paper strives to adopt a critical standpoint towards the provisions of the ethical guidelines in place, arguing that adopting a one-size-fits all approach is not effective due to the specificities of the content distribution industry.

2. "We have to go back" – A brief history of recommendation systems

Recommendation engines are not new inventions, actually they have been around for almost two decades. Knowing how these recommendation systems work can enable us to map the issues that may arise due to their widespread application. It is important to note that this study only gives an outline sketch of the functioning of these systems, it does not strive to give a specific or comprehensive analysis. Descriptions thus may be restricted to general outlines of the concepts discussed, but this level of interpretation should provide the reader with the necessary background to assess the extent of the issues related to AI-powered recommendation systems. At a basic level, recommendation systems consist of machine learning algorithms, which are a subset of AI. "Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention" (URL 6). The machine learning algorithms used to build recommendation systems can be categorized based on the method they use for filtering. Three categories can be identified from the perspective of the filtering method: content-based, collaborative and knowledge-based filtering. The first recommendation systems implemented content-based filtering, while the more advanced ones applied collaborative-filtering methods. Further, the early recommendation engines relied solely on user ratings when making suggestions.

2.1. Content-based filtering

Content-based filtering is a filtering method based, on the one hand, on assigning features/descriptors to every content in the database and, on the other hand, on profiling a user's behaviour using data extracted from the explicit user contributions (rating previously accessed titles or keyword searches). In this model, the user feeds the recommendation system with relevant information, which is used to generate a user profile. The recommendation system assigns items to the list of search results or recommended titles if there is a match between the descriptive attributes of certain media content (movies or series) and the characteristics of the media content used to build the user profile. Aggarwal describes the operation of these systems through providing the example of a user called John who gave a high rating to the movie *Terminator* (Aggarwal 2016, 14). As the descriptors of *Terminator* match the majority of the genre keywords for *Alien* and *Predator*, these movies will be recommended to John (Aggarwal 2016, 14). The following figure illustrates how the system works:

The advantage of content-based recommendation systems is that they do not need a particularly large dataset, as the recommendations are specific to a certain user. These systems are, however, limited, as they only recommend items with similar properties. Sticking to the example of John, if he prefers the genres science fiction, horror and action, he will never be recommended *The Crown*, a historical drama series. This is considered to be a disadvantage as it "tends to reduce the diversity of the recommended items" (Aggarwal 2016, 15).

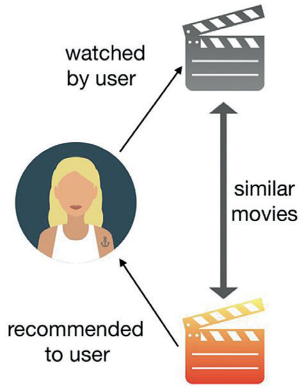

Figure 1. Example of a content-based recommendation system (Source: URL7)

2.2. Collaborative filtering

Collaborative filtering utilizes other users' profiles when making recommendations. The main idea behind these systems is that when users have displayed similar interests and rating patterns in the past, it is likely they are going to have similar preferences in the future. In the 1990s, there were attempts to predict user preferences in order to tailor search result lists and recommendations with collaborative-filtering methods. One of the first recommendation engines was GroupLens, which was used as collaborative-filtering system for Usenet news. A pilot trial was performed which started by inviting users from selected newsgroups to rate pieces of news on a scale of 1–5 (Konstan et al. 2000). Ratings then were used to generate predictions embedded into the recommendations. Later, service providers enhanced their systems, adding more relevant factors to the assessment matrix. For instance, besides active and explicit user contributions (ratings), providers started to incorporate active but implicit contributions to their recommendation engines (such as an assessment of the user's clicking history or watching time).

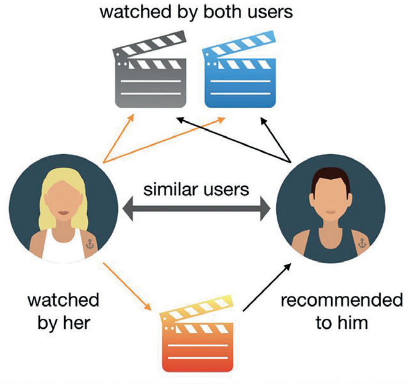

Figure 2. Example of a collaborative recommendation system (Source: URL 7)

Naturally, this is also a simplified description of collaborative filtering, but for the purpose of this study, it is enough. Modern recommendation engines do not use a clean version of the described filtering methods; indeed, the majority of the existing recommendation engines combine elements of different filtering techniques. These recommendation engines are often called hybrid recommendation systems.

3. "The truth is out there" – How AI-powered recommendation engines work in a nutshell

At this point one may ask why are AI-powered recommendation systems the focus of interest all of a sudden, if the base technology, i.e. machine learning algorithms, has been around for decades? Perhaps because AI has only recently reached the level of development that makes their functioning comparable to human thinking and allows them to perform tasks requiring (close-to-) human intelligence. There are many definitions of AI, but there is one common element in all of them: AI should be able to mimic intelligent human behaviour. For the purposes of this study, we use the working definition set out by the European Commission's Communication on AI (URL 8). According to the proposal:

> *"Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications)."*

However, AI is not a uniform technology, in fact there are several categories of AI, depending on their level of independence and scope of functioning. Colloquially, AI is often identified with what Kurzweil called "strong artificial intelligence (SAI)" (Kurzweil 2005). Here, SAIs are considered intelligent agents that are able to perform any intellectual task a human, in other words, it refers to conscious machines with full human cognitive abilities (URL 9). However, it has been stated that when singularity is reached, machine intelligence will exceed human intelligence; thus humans will become unable to understand and control technological development (Kurzweil 2005). Although technology can develop at a frightening speed, SAI yet remains within the domains of science fiction. The majority of AIs currently in use correspond to the notion of narrow artificial intelligence (NAI), as also introduced by Kurzweil (URL 10). NAI systems are only capable of performing specific tasks – albeit with high efficacy – but they lack the cognitive complexity of human thinking.

Due to the exponential development in the areas of algorithm programming, computational power and the availability of massive, easily accessible and transferable data pools, AIs have undergone unprecedented development in the past couple of years. Advances in building deep neural networks have led to the invention of deep learning, which is a subcategory of machine learning. The connection between AI, machine learning and deep learning are shown in the figure below:

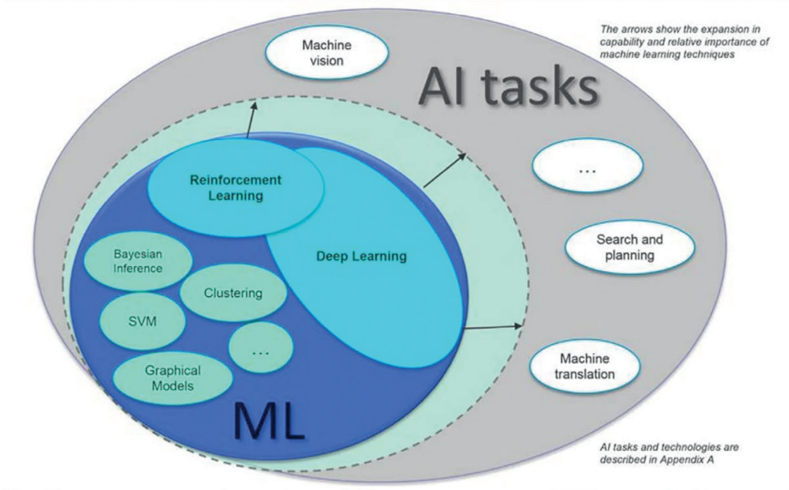

Figure 3. The connections between AI, machine learning and deep learning. Source: (Cambridge Consultants 2019)

Deep learning is a machine learning technique that uses artificial neural networks, which mimic the structure of the human brain (Cambridge Consultants 2019, 20). Deep learning techniques are now applied within the most advanced recommendation systems, like Netflix's.

4. "Winter is coming" – The risks recommendation systems pose to certain fundamental rights

A key aim of this study was to assess the ethical issues related to recommendation systems through the lens of human rights. A human-rights-based approach is relevant, because fundamental rights are core values that are recognized globally and are set out by various international legal instruments. Furthermore, legal instruments adopted within the European Union, for instance, are all rooted in these fundamental rights, and all the institutions of the EU and its Member States are bound to abide by Charter of Fundamental Rights of the EU (2012). The human-rights-based approach is favourable for one more reason: the EU's Ethics Guidelines for Trustworthy AI (AI HLEG 2019) and the proposed AI Code (URL 11) share the view that a human-rights-driven approach is the key to building trustworthy AI. There are some approaches that take into account the human rights dimensions of recommendation systems (but do not focus solely on the fundamental rights aspects of AI-powered recommendation systems). Milano et al. suggest a taxonomy in which recommendation systems are categorized along two dimensions (Milano et al. 2020). The first dimension catalogues the risks identified in connection to recommendation systems based on whether they negatively affect the utility of some stakeholders or constitute a rights violation. The second dimension categorizes risks based on the severity of the impact: some may cause immediate harm, while some only cause an exposure to the future risk of harm. This study focuses solely on the rights dimension of the recommendation system-related risks. It has to be noted though that some overlaps may be observed between some elements of these categories. For example, inaccurate recommendations are normally considered a utility issue, yet if they persistently appear on a larger scale, there is an inherent risk of rights violation (such as unfair treatment). In 2018, controversy revolved around Netflix personalizing movie posters shown to users. The company was accused of personalizing the movie poster selection based on ethnicity, as black viewers were presented posters featuring black cast members (URL 12). What seemed to be at first just a flaw in the algorithmic design turned out to be a risk to the principles of equality and non-discrimination.

Fundamental rights are impacted by AI-powered recommendation systems extensively; the following rights are particularly affected:

- Dignity – human autonomy
- Integrity
- Privacy
- Freedom of information
- Equality, non-discrimination
- Diversity.

Projecting the ethical issues identified through a literature review onto the catalogue of fundamental rights laid down in international legal instruments (especially in the Charter of Fundamental Rights of the European Union), the following risk map was drawn up. However it must be borne in mind that the identified risks intertwine and some of the identified risks affect different rights (especially the lack of transparency).

Affected right	**Risk**
Dignity – human autonomy	Lack of transparency – black box
Integrity	May lead to addiction Inappropriate content
Privacy	User profiling and data leakage Data publishing Algorithm design User interface design Experimentation on user groups
Freedom of information	Filter bubble Lack of transparency – black box
Equality, non-discrimination	Activity bias Algorithmic bias Cognitive bias of the user
Diversity	Lack of transparency – black box with lack of diversity in the recommendations

Table 1. Risk map of AI-powered recommendation systems

4.1. Dignity – human autonomy

Human autonomy when the user browses the system to choose media content to consume is a mirage to some degree, because the options recommended by the algorithm are filtered media content, deemed to be relevant for the individual user by the recommendation system. Such filtering limits the list of available options, thereby in reality curbing the users' freedom of choice.

The functioning of recommendation systems remains hidden from the user, with such tools operating in the background, unnoticed as part of the user experience. Algorithmic decision-making tools – not limited to recommendation systems – in fact work as a 'black box' (Pasquale 2015) system, where users cannot tell why and how the system generated one specific output. One reason behind this is the fact that we are talking about sophisticated program codes and complex mathematics often guarded as trade secrets of the provider. One may argue that the release of the source codes does not really carry relevant information to the general public. The European Union Agency for Fundamental Rights' (FRA) report highlights that one aspect of preserving human dignity is to inform people about the use of AI, enabling them to provider informed consent (FRA 2020, 60). Transparency in the case of AI thus equals explicability: Knowledge of the principles of functioning and the factors that were taken into consideration when compiling the list of recommended content that would be enough to contribute to user awareness and to allow them to reach informed consent.

4.2. Integrity

According to the Charter of the Fundamental Rights of the EU, the right to integrity of the person means respect for one's physical and mental integrity. Research shows that the use of recommender systems paired with psychological factors, such as a lack of self-control or lack of self-esteem complemented with certain motives (the motive of information seeking) can lead to excessive usage (Hasan et al. 2018). Excessive internet usage and content consumption are known can have negative impacts on individuals' psychosocial well-being (Young 2004.), and can have negative consequences on individuals, such as emotional problems, relational problems, sleep-difficulties and performance problems (Andreassen 2015). Recommender systems that are designed to manipulate users – sometimes with subliminal techniques – thus may adversely affect the users' mental integrity.

Inappropriate content (Milano et al. 2020) may also have negative impacts on a person's integrity, although there is no common understanding what the term 'inappropriate content' means. Inappropriate could mean content that is erroneously suggested, contradicting a user's preferences and predictions reached by analysing the factors assessed by the recommendation system. Inappropriate can also mean that recommendations are not culturally appropriate for the individual users or certain user groups (Souali et al. 2011). This was the case when the Christian community of Brazil petitioned to remove the movie titled *The First Temptation of Christ* from Netflix's catalogue because it portrayed Jesus as a homosexual figure (URL 13). Inappropriate can be also interpreted in terms of certain vulnerable audience groups, such as minors. The popular American teen drama titled *13 Reasons Why* was severely criticized for romanticizing suicide, yet it appeared on the lists of recommendations of teenage users. The show is very popular among adolescents, although it could pose a serious risk for mentally unstable people, or people with mental health issues (URL 14). There are rules in the European audiovisual media regulation which aim to restrict the free flow of content that is detrimental to minors. For instance, content that can cause serious harm to the physical, moral or mental development of children should be restricted to adult audiences. This is an obligatory provision of the AVMS Directive since 2010 and applies to on-demand streaming providers as well, pursuant to which they must tune their recommendation systems to take into consideration the user's age and the parental control settings.

4.3. Privacy

Privacy is one of the key challenges identified and the protection of personal data is the most cited fundamental right in the AI-related discourse. The right to privacy has paramount importance in the case of recommendation systems, which profile users to fine-tune content recommendations. The factors that are taken into consideration by recommendation systems – introduced in Section 1. – are designed with regard to the availability of user data. There are five problematic areas that were identified in relation to recommendation systems by Paraschakis, three of which (1. 2. and 5.) are privacy related (Paraschakis 2017):

1. user profiling and data leakage
2. data publishing
3. algorithm design
4. user interface design
5. experimentation on user groups.

Paraschakis draws attention to the fact that behavioural profiling is often done without acquiring informed consent, as privacy notices hidden behind hyperlinks that follow "I consent" checkboxes often remain unread by the user (Paraschakis 2017). Unsolicited data collection is also common, because user profiles are generally enhanced with data obtained from external sources, such as cookies, social networks or information brokers, despite the fact that the integration of external sources can lead to vulnerability and could lead to data breaches. Friedman et al. consider the actions of external adversaries who attempt to de-anonymize data one of the biggest privacy-related risks to recommendation systems (Friedman et al. 2015). Paraschakis adds that companies often release large datasets from their services that contain private data. Although personally identifiable information (such as names and email addresses) are anonymized, there are quasi-identifiers (birth date, gender, location) that can be combined to identify users. Companies also often test new versions of their recommendation systems on randomly selected user groups. The most popular method for testing new algorithms is A/B testing, in which two variants of the same webpage is shown to different user groups (a control and treatment group). A famous example of such an experiment is research conducted through Facebook as part of an emotion experiment in 2014 (Kramer et al. 2014). In the experiment, members of different user groups were shown differently curated news feeds. Users who were shown more positive posts reported feeling happier, while the people who had seen negative images more frequently felt unhappy and showed signs of depression. Besides the issue of the unethicalness of the experiment, the research also drew attention to the lack of informed consent of the users, who weren't informed prior to the experiment that they were part of such a research, or of the handling of their personal data for the purposes of the research.

4.4. Freedom of information, freedom of expression

Freedom of expression involves freedom of information and means the right to receive and impart information and ideas without interference. Yet, the main objective of recommendation systems is to interfere with the flow of information by selecting relevant information (or information considered relevant) for each individual user, with the aim to enhance the user experience and promote engagement. By consuming overly-personalized media content, users can easily become isolated from media content that is outside their comfort zone and ideas different from their own ideology, resulting in them getting stuck in cultural and ideological bubbles, named filter bubbles (Pariser 2011.) The user may not necessarily notice getting into such a filter bubble due to the elaborate design of the filtering system and the lack of transparency.

It is important to examine the operation of recommender systems from the perspective of the content creators (artists) as well. Streaming services are generally good platforms for emerging creators and narrow-niche genres, because streaming service providers facilitate the worldwide (or regional) distribution of audiovisual works of any genres, and as audiences can find media content that is often not available through traditional distribution. Due to the long-tail effect (Anderson 2006), offering niche content is profitable for streaming service providers. However, these providers are also in a gatekeeper position, in that they have a direct impact on the media content they make available to the public (Koltay 2019, 82). This means that streaming service providers can arbitrarily control what gets popular and what gets lost among the myriad of content, reducing the visibility of certain creators and widening the gap between well-known global studios and smaller studios making art films or niche movies. We do know that some creators sign deals with streaming providers to produce, to distribute and even to feature their works, resulting in a further imbalance between the larger and richer and well-known outlets and the small studios. In this system, service providers have little to no regard to the credibility of the information conveyed, blurring the boundaries between fictional works and documentaries, which carries the risk that, due to the fact that they can reach millions of people, they can amplify the spread of false or misleading information. Netflix received harsh criticism for signing a deal with Gwyneth Paltrow's Goop, which is known to feature a pseudoscientific lifestyle documentary (URL 15), but was also criticized for producing a similar documentary series titled *Down to Earth* with Zac Efron (URL 16), and for the documentary *Seaspriacy*, which was accused of containing misleading claims about commercial fishing (URL 17).

4.5. Equality, non-discrimination

The Charter of Fundamental Rights of the EU (2012) considers all people equal and sets out that any discrimination based on protected characteristics – such as sex, race, colour, ethnic or social origin, genetic features, language, religion shall be prohibited– (Article 21). Furthermore, it sets out that the EU shall respect cultural, religious and linguistic diversity (Article 22). However, recommendation systems are biased by design, as they draw up patterns, which they then use to generalize users. The FRA's report also draws attention to the fact that the "very purpose of machine learning algorithms is to categorize, classify and separate" (FRA 2020, 68). Baeza-Yates differentiates between three types of biases that characterize recommendation systems and distort the list of recommended media content (Baeza-Yates 2020):

1. Activity bias, which refers to the distorting effects of the attributes that are automatically assigned to users upon browsing and searching, such as gender, age, location, language of the service.

2. Algorithmic bias, which refers to the distortion that can be traced back to the programming of the algorithm. Recommendation systems work with sets of variables, weigh each factor and rank each property differently, where the principles of weighing and ranking are coded into the system by programmers having their own biases. One form of algorithmic bias is observation bias (Farnandi et al. 2018), which refers to the feedback loops generated to specific groups of users. The term "feedback loop" has been used in software development for some time now and it refers to a situation where the outputs of a system are loaded back to be used as inputs. Using the outputs generated by the system as teaching data amplifies bias and leads to the development of filter bubbles (Mansoury et al. 2020; Jiang et al. 2019). Observation bias is also caused by population imbalance (Farnandi et al. 2018), whereby existing social patterns are reflected in the system's decisions. Observation bias can mostly occur during the application of collaborative-filtering tools, which rely on interpersonal relationships (other, similar people's preferences) to filter information (Bozdag 2013). Bozdag also mentions popularity bias – which means popular content often gets highlights and thus gets even more popular – hindering the diversity of recommendations.

3. The cognitive biases including confirmation bias and other behavioural biases of the user, which also affect the functioning of the recommender system, as the user is able to tune and teach the system through his or her choices. Unconscious decisions are taken into consideration by the recommender system, leading to the formulation previously discussed filter bubbles. According to Pariser, confirmation bias is often enhanced by personalization algorithms, because consuming information or media content that conforms to one's taste and ideology causes pleasure (Pariser 2011), while diverse opinions and content can lead to cognitive dissonance.

4.6. Diversity

The entertainment industry has often been described as an industry fuelling blockbuster culture. (Anderson 2006.) The term 'blockbuster' has been used for movies since the 1970s (one of the first blockbusters was Steven Spielberg's *Jaws*) and refers to fast-paced and exciting movies, that tend to generate interest beyond the cinema (Shone 2004) and are capable of reaching an extremely wide audience (see the Marvel Cinematic Universe). Although there is a vast amount of available audiovisual works, there are only a limited number of blockbusters. AI-powered recommendation systems may have the effect that they "reinforce the popularity of already popular products" (Fleder et al. 2009, 679), as they are more likely to appear on the top of the list of recommendations and in the list of many difference audience categories. However, it is also argued that these systems can enable members of the audience to find niche content (Anderson 2006).

It is worth highlighting that the promotion of European audiovisual culture with regulatory tools is not unbeknownst in EU law, and in fact it is an obligation already present in the Audiovisual Media Services Directive (AVMSD). The AVMSD explicitly obliges on-demand service providers (such as Netflix and other streaming service providers) to foster the European audiovisual culture and movie industry by reserving at least a 30% share of European works in their catalogues (Article 13). Additionally, the providers also have to ensure the prominence of those works.

Mehrota et al. (2018) point out that modern recommendation systems serve two-sided markets, so algorithms must be optimized in a way to take the interest of the supply side (artists) into consideration as well.

5. "To boldly go…" – Ethical codes as tools to tackle the challenges of AI

The risks of AI have been recognized by several international organizations. In recent years, several AI ethical codes have been drafted to mitigate these risks. Notable examples are the following:

- The Ethics Guidelines for Trustworthy AI of the European Commission's High-level Expert Group (AI HLEG 2019) on Artificial Intelligence;

- The OECD's Recommendation of the Council on Artificial Intelligence (OECD 2019);
- The Beijing AI Principles drafted by the Beijing Academy of Artificial Intelligence (BAAI 2019); and
- Guidelines adopted by the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems titled "Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems" (IEEE 2019).

This paper discusses two of these guidelines in detail: HLEG's Ethics Guidelines and the OECD's Recommendation as these instruments are the most relevant from the perspective of the EU's legal framework. Additionally, it has to be noted that all the ethical codes are surprisingly similar to each other as they grasp AI from the same perspective, whereby they place values such as trustworthiness and fairness at the centre. This section not only gives a detailed description of these guidelines, but specifically highlights which provisions are relevant in terms or recommendation systems and why. The second part of this section then provides constructive criticism of the ethical codes discussed.

5.1. How do ethical codes drive the future of AI?

The OECD's Recommendation of the Council on Artificial Intelligence sets out 5 value-based principles for a responsible stewardship of trustworthy AI. The principles for AI are that it should operate in line with:

- The pursuit of inclusive growth, sustainable development and the well-being of humankind.
- Human-centred values and fairness. This principle includes due consideration for the rule of law, human rights and democratic values. In order to abide by this principle, specific mechanisms and safeguards should be implemented into AI systems, such as a capacity for human determination.
- Transparency and explicability.
- Robustness, security and safety.
- Accountability.

These five principles are targeted towards those driving the development of AI, such as those who design and operate systems. To complement these five principles, complementary recommendations were added for policy-makers to take into consideration.

The HLEG's guidelines characterize trustworthy AI as meaning lawful, ethical and robust. The EU's AI ethics principles are rooted in the respect for fundamental rights, and thus the ethical principles set out by the guidelines are based on tangible rights set out by existing international legal instruments. The guidelines list 4 ethical principles, terming them ethical imperatives:

1. respect for human autonomy
2. prevention of harm
3. fairness
4. explicability.

Respect for human autonomy means that humans must be able to maintain self-determination over themselves and are entitled to be protected from manipulation, coercion, deception and conditioning: AI systems thus should refrain from applying techniques that manipulate human beings and should be designed to rather augment, complement and empower human skills. The principle of prevention of harm means that AI systems should be designed to protect human dignity, as well as mental and physical integrity, which is basically the ethical implementation of the first law of robotics, made famous by Isaac Asimov in his work of science fiction *I, Robot* (Asimov 1950). The principle of fairness means – among others – that AI systems should ensure that individuals and groups are free from unfair bias, discrimination and stigmatization. Explicability means that the processes should be transparent and an explanation should be provided why the system reached a particular decision, in order to build and maintain the user's trust.

The issues that were in identified in relation to how recommendation systems affect certain fundamental rights are also risks to the principles set out by the ethics codes. The Ethics Guidelines for Trustworthy AI sets up a non-exhaustive list of requirements to achieve trustworthy AI, which are all relevant to mitigate the risks of AI-powered recommendation systems to fundamental rights as long as the industry players are willing to align their behaviour to them. The list of requirements consists of requirements such as human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity. However as argued in the next subsection ethical principles and requirements do not create domain-specific obligations and are not enforceable thus they are not sufficient tools to alleviate concerns.

5.2. Critical remarks concerning the ethical codes

A criticism that is often voiced towards codes of ethics in general can be projected to AI specific codes as well; namely that ethics codes tend to be the results of long negotiations between industry members and the industry members and state actors, where the end text is almost always a result of many compromises. This tends to lead to a set of diluted norms, stripped down to the most important core values and principles recognized by all parties. Setting common values are important elements of marking out a normative framework, but they are not very effective if the norms that should fill the frames are missing.

Given their nature, ethical principles tend to be worded vaguely, making the norms set out overbroad. Here, the values and principles set out by the AI ethical codes are important elements of a normative framework, but not exactly AI specific. Héder notes that the provisions set out in AI ethics guidelines can be applied to any novel technology; for example, if one were to change the term "AI" to the term "water boiler", the guidelines would still be interpretable (Héder 2020). Indeed, the majority of the discussed principles have been set out already in different scientific areas, such as bioethics. There is also the issue of general documents not being able to react to domain-specific issues in an appropriate manner. NAIs are used in several sectors nowadays, ranging from application in autonomous vehicles to facial recognition. Although they have common elements that can be addressed by horizontal codes of ethics, they also pose special challenges that are not dealt with by general ethics codes. For example, freedom of information is a right that is significantly impacted by AI-powered recommendation systems, meanwhile it is almost completely irrelevant in the case of autonomous weapons systems, in which case the right to life has more relevance. Current ethical codes are thus unable to tackle those issues that stem from the specific functions that AIs applied in specific areas have, thus suggesting regulatory blind spots exist.

As these codes are often too vague, they are unable to influence the signatories' behaviour when it comes to real-life application. As these are soft-law instruments and do not contain tangible obligations, abiding by the norms and the manner in which compliance is realized is largely dependent on the willingness of the members of the industry. The lack of mechanisms for creating compliance is a particular weak point of ethics codes according to many scholars (Hagendorff 2020; Larsson 2020). Without setting out mandatory rules to oblige parties to a certain conduct or to refrain from certain behaviours, there is no way to impose sanctions on those who fail to act in accordance with the principles. The infringement of ethical codes may also result in disadvantage (the disapproval of society, loss of clients, etc.), but these differ from the sanctioning framework set out by legal norms. Larsson's main reason for concern is that it is unclear what the relationship is between ethics codes and pieces of legislation and notes that ethical guidelines are essentially being drafted by industry players with the incentive to avoid stronger state-regulation (Larsson 2020).

6. Concluding remarks – How can recommendation systems promote culture and diversity?

To effectively tackle the domain-specific issues of AI-powered applications and to liquidate the regulatory blind spots identified in Section 5, a more detailed set of norms should be drafted for each individual application domain. Horizontal AI guidelines are too general to tackle the issues that require sectoral tailor-made solutions. In the scientific literature, there have been several suggestions made to address the risks of AI-powered recommendation systems such as those identified in Section 4., which should provide a good starting point for further regulatory initiatives.

Increasing the transparency would be one cure to the majority of the problems discussed, because making recommendation systems more transparent would positively contribute to human autonomy, diversity and privacy. The developers of AI should be mandated to explain how recommendation systems work, what factors they are considering to generate the outputs and how they handle personal data to generate recommendations. Héder warns that if transparency is not defined at an appropriate level, AI development may be hindered (Héder 2020). Furthermore, he argues that transparency has to be combined with some measure of intelligibility to avoid "pretend transparency". Educating users to move safely and comfortably in the digital world is not a novel thought. The concept of digital literacy was developed as long ago as the 1990s and refers to "the ability to understand and use information in multiple formats from a wide variety of sources when it is presented via computers" (Gilster 1997). Media literacy – which is a narrower concept – refers to those skills that allow users to use the media; for example, to access information and to critically assess media content. Along these concepts, there is a need to introduce the concept of algorithmic literacy (URL 18). Algorithmic literacy should involve the skills to critically assess the recommendations made by algorithms, to exploit the recommendation systems in a manner that serves the best interest of the individual and to develop an awareness of algorithmic biases and how to avoid their influence.

Ensuring a diversity of content is a factor that should be incorporated into algorithms (Castells et al. 2015). The recitals of the AVMSD (recital 35) propose the need to ensure prominence by labelling metadata, facilitating access and setting up dedicated sections in catalogues. The AVMSD does not explicitly mention recommendation systems as a tool to enhance visibility, but does mention that fine-tuning the algorithms to recommend European content can also be considered a viable option to increase the reach of European audiovisual works. Optimizing recommendation systems to promote audiovisual culture and the interests of the supply side would definitely mean a shift from subordinating company policies to audience engagement, and in the long term these measures could contribute to a more diverse media landscape. As members of the audience would be able to find European and national works along with the works of emerging artists, the fine-tuning of recommendation systems could contribute to the promotion of cultural exploration.

Besides enhancing the algorithmic designs, the previously mentioned risks can be mitigated by giving the users more freedom to customize their experience. In the area of privacy, explicit privacy controls should be incorporated into the systems, including allowing the users to decide which data is to be shared and with whom (Paraschakis 2017). Customizable settings should be introduced in filtering as well, because if one can choose to filter adult content (which is what parental control tools are about), one should also be able to select their preferences in terms of other factors (such as genres, actors, directors). More options to customize the service would also contribute to enhancing the autonomy of the user.

References

Aggarwal, C.C. *Recommender Systems: The Textbook*. Springer, 2016.

Anderson, C. *The Long Tail*. Hyperion: New York, 2006.

Andreassen, C.S. "Online Social Network Site Addiction: A Comprehensive Review." *Current Addiction Reports*, 2 (2) (2015): 175–184.

Asimov, I. *I, Robot*. Gnome Press, 1950.

Baeza-Yates, R. "Bias in Search and Recommender Systems." *RecSys '20: Fourteenth ACM Conference on Recommender Systems*, September 2020. https://doi.org/10.1145/3383313.3418435

Beijing Academy of Artificial Intelligence. *Beijing AI Principles*. 2019.

Bozdag, E. "Bias in algorithmic filtering and personalization." Ethics and Information Technology 15, (2013): 209–227. https://doi.org/10.1007/s10676-013-9321-6

Cambridge Consultants: Use of AI in Online Content Moderation 2019 Report Produced on Behalf of Ofcom. 2019.

Castells P., Hurley N.J., Vargas S. "Novelty and Diversity in Recommender Systems." In: Ricci F., Rokach L., Shapira B. (eds) *Recommender Systems Handbook*. Springer, Boston, MA., 2015. https://doi.org/10.1007/978-1-4899-7637-6_26

Fleder, D., & Hosanagar, K. "Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity." *Management Science*, 55 (5) (2009): 697–712. http://dx.doi.org/10.1287/mnsc.1080.0974 "

Forte, G., Favieri, F., Tedeschi, D., Casagrande, M. "Binge-Watching: Development and Validation of the Binge-Watching Addiction Questionnaire." Behavioral Sciences. 11, no. 2: 27. (2021) https://doi.org/10.3390/bs11020027

Friedman, A., Knijnenburg. B., Vanhecke, K., Martens, L., Berkovsky, S. "Privacy aspects of recommender systems." In: *Recommender Systems Handbook*, 649-688 New York: Springer Science+Business Media, 2015.

Gilster, P. *Digital Literacy*. John Wiley & Sons, 1997.

Hagendorff, T. "The Ethics of AI Ethics: An Evaluation of Guidelines." *Minds & Machines* 30, (2020): 99–120. https://doi.org/10.1007/s11023-020-09517-8

Hasan, M.R., Jha, A.K., and Liu, Y. "Excessive use of online video streaming services: Impact of recommender system use, psychological factors and motives." *Computers in Human Behaviour*, 80 (2018):220–228.

Héder, M. "A Criticism of AI Ethics Guidelines." *Információs Társadalom* 20, no 4. (2020) https://doi.org/10.22503/inftars.XX.2020.4.5

High-Level Expert Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI. (2019)

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems (2019).

Ray Jiang, Silvia Chiappa, Tor Lattimore, András György and Pushmeet Kohli. "Degenerate Feedback Loops in Recommender Systems." In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, Honolulu, USA. 2019.

Koltay, A. *New Media and Freedom of Expression: Rethinking the Constitutional Foundations of the Public Sphere*. Hart Publishing, 2019.

Konstan, J.A. et al. "GroupLens: Applying collaborative filtering to Usenet news." In *February 2000 Communications of the ACM* 40(3) (2000): 78.
https://doi.org/10.1145/245108.245126

Kramer, A., Guillory, JE., Hancock, JT. "Experimental evidence of massive-scale emotional contagion through social networks." *Proceedings of the National Academy of Sciences*, 111 (24) (2014):8788–8790.
https://doi.org/10.1073/pnas.1320040111

Kurzweil, Ray (2005), *The Singularity is Near.* Viking Press, 260.

Larsson, S. "On the Governance of Artificial Intelligence through Ethics Guidelines." *Asian Journal of Law and Society*, 7(3) (2020):437–451.
https://doi.org/10.1017/als.2020.19

Masoud Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher B. and Burke R. "Feedback Loop and Bias Amplification in Recommender Systems." In *CIKM '20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2145–2148. 2020.
https://doi.org/10.1145/3340531.3412152

Mehrota, R., McInerney, J., Bouchard, H., Lalmas, M. and Diaz, F. "Towards a Fair Marketplace: Counterfactual Evaluation of the trade-off between Relevance, Fairness & Satisfaction in Recommendation Systems." In *CIKM '18: Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2243–2251. 2018.
https://doi.org/10.1145/3269206.3272027

Milano S., Taddeo, M. and Floridi L. "Recommender Systems and their Ethical Challenges." *AI & Society*, 30 (2020):957–967.

OECD. *OECD's Recommendation of the Council on Artificial Intelligence.* OECD/LEGAL/0449. 2019.

Paraschakis, D. "Towards an ethical recommendation framework." In *Conference Proceedings 11th IEEE International Conference on Research Challenges in Information Science*, 211–220, 2017.
https://doi.org/10.1109/RCIS.2017.7956539

Pariser, E. *The filter bubble: What the internet is hiding from you*. London: Penguin Press, 2011.

Pasquale, F. *The Black Box Society: The Secret Algorithms That Control Money and Information.* Harvard University Press, 2015.

Riddle, K., Peebles, A., Davis, C., Xu, F., & Schroeder, E. "The addictive potential of television binge watching: Comparing intentional and unintentional binges." *Psychology of Popular Media Culture*, 7(4) (2018):589–604.

Shone, T. *Blockbuster*. London: Simon & Schuster UK, 2004. 27–40.

Souali, K., El Afia, A. and Faizi, R. "An automatic ethical-based recommender system for e-commerce." In *2011 International Conference on Multimedia Computing and Systems*, 1–4, 2011.
https://doi.org/10.1109/ICMCS.2011.5945631

Young, K. "Internet Addiction: A New Clinical Phenomenon and Its Consequences." *American Behavioural Scientist*, 48 (2004):402–415.
https://doi.org/10.1177/0002764204270278

URL 1: McFadden, C. "The Fascinating History of Netflix." Interesting Engineering, Accessed 17. May 2021.
https://interestingengineering.com/the-fascinating-history-of-netflix

URL 2: Dean, B. "Netflix Subscriber and Growth Statistics: How Many People Watch Netflix in 2021?" Backlinko, Accessed 17 May 2021.
https://backlinko.com/netflix-users

URL 3: McCluskey, B. "9 Netflix slang terms besides 'Netflix and Chill' every binge watcher needs to know." Business Insider, Accessed 17. May 2021.
https://www.businessinsider.com/netflix-slang-and-their-meanings-2017-3

URL 4: Merriam-Webster Dictionary. Accessed 17. May 2021.
https://www.merriam-webster.com/dictionary/binge-watch

URL 5: West, K. "Unsurprising: Netflix Survey Indicates People Like To Binge-Watch TV." Cinema Blend, Accessed 17. May 2021.
https://www.cinemablend.com/television/Unsurprising-Netflix-Survey-Indicates-People-Like-Binge-Watch-TV-61045.html

URL 6: SAS. "Machine learning" Accessed 17 May 2021.
https://www.sas.com/en_us/insights/analytics/machine-learning.html#:~:text=Machine%20learning%20is%20a%20method,decisions%20with%20minimal%20human%20intervention.

URL 7: Grimaldi, E. "How to build a content-based movie recommender system with Natural Language Processing." Towards Data Science, Accessed 17. May 2021.
https://towardsdatascience.com/how-to-build-from-scratch-a-content-based-movie-recommender-with-natural-language-processing-25ad400eb243

URL 8: Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe, Brussels 25.4.2018 COM(2018) 237 final. Accessed 17 May 2021.
https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN

URL 9: Marr, B. "What Is The Difference Between Weak (Narrow) And Strong (General) Artificial Intelligence (AI)?" LinkedIn, Accessed 17 May 2021.
https://www.linkedin.com/pulse/what-difference-between-weak-narrow-strong-general-artificial-marr/?trackingId=PeGWhIxST8m69dhMG4pACA%3D%3D

URL 10: Kurzweil, R. "Long Live AI." Forbes, Accessed 17. May 2021. https://www.forbes.com/forbes/2005/0815/030.html?sh=426123fb7e8f

URL 11: Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts COM/2021/206 final, Accessed 26. August 2021.
https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206

URL 12: Iqbal, N. "Film fans see red over Netflix 'targeted' posters for black viewers." The Guardian, Accessed 17. May 2021.
https://www.theguardian.com/media/2018/oct/20/netflix-film-black-viewers-personalised-marketing-target

URL 13: Reuters. “Netflix show portraying Jesus as gay sparks anger in Brazil.” The Guardian, Accessed 17 May 2021 https://www.theguardian.com/world/2019/dec/17/netflix-show-portraying-jesus-as-gay-sparks-anger-in-brazil

URL 14: MollyKate, C. “Why ‘13 Reasons Why’ Can Be Triggering for People Coping With Mental Illness.” Teen Vogue, Accessed 17 May 2021. https://www.teenvogue.com/story/13-reasons-why-can-be-triggering-coping-with-mental-illness

URL 15: Mahdawi, A. “Goop has a Netflix deal – this is a dangerous win for pseudoscience.” The Guardian, Accessed 17. May 2021. https://www.theguardian.com/commentisfree/2019/feb/08/goops-deal-with-netflix-is-a-dangerous-win-for-pseudoscience

URL 16: Dodgson, L. ‘All the problematic pseudoscience shared by Zac Efron’s health guru and guests in his new Netflix show ‘Down to Earth’.” Insider, Accessed 17 May 2021. https://www.insider.com/pseudoscience-in-zac-efron-new-netflix-show-down-to-earth-2020-7

URL 17: McVeigh, K. “Seaspiracy: Netflix documentary accused of misrepresentation by participants.” The Guardian, Accessed 17. May 2021. https://www.theguardian.com/environment/2021/mar/31/seaspiracy-netflix-documentary-accused-of-misrepresentation-by-participants

URL 18: Pew Research Center. “Theme 7: The need grows for algorithmic literacy, transparency and oversight.” Accessed 20. May 2021. https://www.pewresearch.org/internet/2017/02/08/theme-7-the-need-grows-for-algorithmic-literacy-transparency-and-oversight/