

# Addictive Algorithms: Setting the Ethical Boundaries of Recommendation Algorithms on Social Media

Cis van Aken (7295758)<sup>1</sup>, Hugo Kuipers (6858309)<sup>1</sup>, Olivier Oude Booijink  
(8365997)<sup>1</sup>, and Jan Stelmaszczyk (6567274)<sup>1</sup>

Utrecht University, Information and Computing Science department, The Netherlands

**Abstract.** This study examines the ethical implications of using recommendation algorithms in content-based social media and identifies gaps in existing regulation. Modern social media platforms attempt to capture users' attention and keep them engaged for as long as possible. For this, they employ several tricks and techniques. One of those being recommendation algorithms to serve tailored content that the user is likely to interact with. In an attempt to figure out what can be regulated about such systems, we conducted a literature review, through which we identified a list of effects that recommendation systems have. Using these effects, we then reasoned on the ethicality of such systems, to ultimately discuss potential regulations if a particular effect was found to be within the company's responsibilities. Using these methods, we found that most negative effects coming from the use of tailored content, fall within the responsibilities of social media platforms to limit. Moreover, we found that current regulation barely exists, leaving plenty of room to fill gaps, for which we provide several recommendations.

**Keywords:** Social Media · Recommendation Algorithms · Ethical Guidelines.

## 1 Introduction

Designing for attention is ubiquitous in social media. Social media platforms utilise many addictive features to increase user engagement [23,12], as this drives the profitability of these platforms. These features have been found to foster addictive behaviours in users [23,2]. The tension between maximising engagement, while avoiding addictive behaviours has been coined the “addiction-engagement” dilemma by researchers [45,44].

Defining social media platforms is not a straightforward task. Numerous authors have attempted to narrow the scope of social media [42,10], but a comprehensive definition that captures all aspects remains challenging. To address this, we will limit the scope of this paper to content-based social media platforms. We define content-based platforms as platforms where users can upload/share videos, images, and text to a (semi-)public feed.

This paper aims to contribute to the ongoing discourse on the ethics and corporate responsibility of social media platforms with regard to their recommendation algorithms. Our goal is to structurally examine the responsibilities of recommendation algorithms and provide advice on how to solve the aforementioned "addiction-engagement" dilemma. Though we apply the framework to a specific context, the created framework could be applied to any IT system to examine its effects and determine responsibilities.

To achieve this, we first provide technical background on recommendation algorithms in section 2. Section 3 elaborates on our methodology and research questions. Section 4 examines the effects of recommendation algorithms on social media platforms and creates an architecture decision map [20] to capture these effects. Section 2 critically examines the ethical responsibilities of social media platforms regarding the stated effects of shareholder, stakeholder and social contract theory as per [3]. Section 6 uses the determined responsibilities to postulate concrete policies for social media platforms with regard to recommendation algorithms. Our findings are discussed in sections 7 and 8 and the conclusions are specified in section 9.

## 2 Background

### 2.1 Algorithms

In essence, algorithms are coded computer instructions that tell a computer system how to operate. In the domain of social media, algorithms serve multiple roles [9]: content ranking, content recommendation, quality screening, unwanted content detection, and advertising.

This research mainly focuses on recommendation algorithms. A recommendation algorithm recommends/displays new, unsubscribed content-based on a user's interests. This is often deducted from viewing history, likes, comments, forwards, and other user-generated data.

A recommendation algorithm is known to determine the content-based-platform's success [46]. For this reason, the precise construction of each recommendation system is often kept secret, although the basic principles are known. Users view content and sometimes interact with this content. Any of these interactions and view cases are stored. This can include likes, watch time, comments, content sharing and many other data points. By labelling all content and measuring user interaction, users can be grouped based on their interests and be shown adequate personalised content. These recommendation algorithms keep improving as each user interaction, or the lack thereof, introduces a new data point to further fine-tune to an individual's interests.

These recommendation algorithms have become very successful at engaging its users, in many cases leading to user addiction to a certain extent [18]. The user is served instant gratification through exploring new content, with minimal effort, as the recommendation algorithm operates so well.

### 3 Method

Through our research, we explore the ethics surrounding recommendation algorithms and figure out to what extent they can and should be regulated. The approach is divided into 3 separate sub-questions, that together reach an answer to our full research question: What ethical responsibilities do social media platforms hold regarding the use of recommendation algorithms and what boundaries can be set?

- Research question 1: What are the societal and personal effects of using recommendation algorithms to serve users content on social media platforms?
- Research question 2: What are the ethical responsibilities of social media platforms when applying recommendation algorithms?
- Research question 3: What forms of regulation could be set to influence the effects of recommendation algorithms?

#### 3.1 Method research question 1

To answer the first research question, and come to a list of effects we could use to reason with further, a literature review was conducted. Prompting when searching was not restricted, however the focus was around terms such as “recommendation algorithms”, “addiction” and “social media”. The goal was to cast a wide net and find a wide variety of different effects, therefore the review was conducted rather freely. When narrowing down a list, one of the criteria we selected on for the reliability of papers, was the amount of times a paper was cited, and if the effect appeared in more than one paper.

The literature review turned into a long, semi-structured text on all these different effects. To make sense of how each effect is related and structure them further, we applied an architecture decision map to the effects. Following the method set out in a different paper, all effects were ordered into categories (personal, societal, economic and technological) and scale (immediate, enabling and structural) [20]. Through this, we ended up with a large model, which was then scoped down to fit our research again. The scoped model is provided and explained in section 4, with the full model available in appendix A. This scoped model provides us with the list of effects that we can then use to reason with when attempting to answer RQ2.

#### 3.2 Method research question 2

In order to get to a workable list of responsibilities that we could potentially discuss regulations on, a different framework is needed. Such a framework was described in another paper and incorporated in our own research. That paper discusses a framework using three different theories of reasoning, to argue what is ethical and what is not for a corporation. We applied these three theories of reasoning to our list of scoped effects and argued if it falls within the scope of responsibilities of a given corporation. for a more elaborate explanation of how the framework works, we suggest reading the source material [3].

### 3.3 Method research question 3

At this stage, we have narrowed down our literature to a table of effects that fall within a corporation's ethical responsibilities. First, we conducted a review on what current regulations are in place, regarding recommendation algorithms specifically. For this, regulations on content restrictions and moderation, not necessarily focusing on regulating algorithms, were left out. Skimming through the websites of regulative bodies such as the European Union is enough to get a good grasp of what is out there. Based on that, we identified gaps in the existing legislature and provided our own recommendations on how those gaps could be filled.

## 4 Effects

There are many effects associated with the use of recommendation algorithms in social media. Of these effects, they often influence one another, either strengthening or weakening them. In an attempt to structure these effects, we have organised them in an architecture decision map [20], which is further elaborated upon in this section.

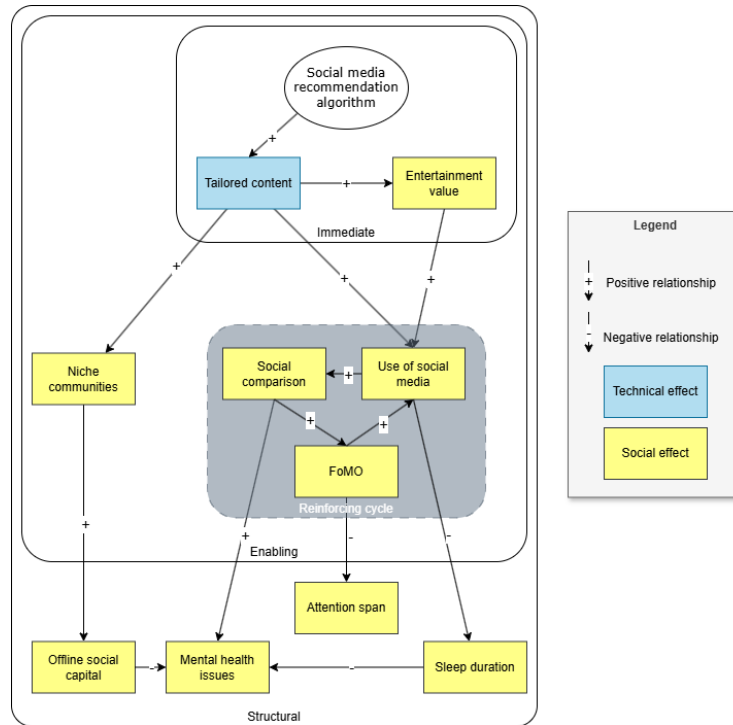


Fig. 1. Scoped architecture decision map

A full architecture decision map is included in Appendix A (Fig. 2). For this section, we limit the scope to the identified personal social effects that flow out of the *tailored content*. All of the relations between these scoped effects are elaborated upon one by one, starting with the enabling level (micro) to finally move onto the structural level (macro). As the effects on the immediate level are given, these are not elaborated upon. Whenever an effect is mentioned, it is written in *italics*.

It is important to note that *use of social media*, *social comparison*, *FoMO*, and create a reinforcing cycle, which perpetuates the effects that indirectly follow from them.

#### 4.1 Enabling effects

The enabling effects refer to effects which arise over time.

**Tailored content  $\xrightarrow{+}$  Niche communities** Researchers have found that online communication tools such as social networks enable people to expand their social networks and become part of new communities that they otherwise would not have, thus finding *niche communities* through *tailored content* [11,32].

**Tailored content  $\xrightarrow{+}$  Use of social media** As content is tailored to users, they spend more time on social media [37]. This relation is partially mediated by the entertainment value, as tailored content tends to be more entertaining, and thus it increases engagement.

**Use of social media  $\xrightarrow{+}$  Social comparison** Social media platforms facilitate a culture of *social comparison* [35], wherein users evaluate themselves against the curated lives of others shown on social media.

**Social comparison  $\xrightarrow{+}$  FoMO** *FoMO* is characterised as "[...] a pervasive apprehension that others might be having rewarding experiences from which one is absent" [26]. Research has found *FoMO* to be fostered by the use of social media [34,27,4].

**FoMO  $\xrightarrow{+}$  Use of social media** Fear of missing out (*FoMO*) has been demonstrated to be a significant factor in determining how often people use social media [26]. Note how this combined with the previous two effects creates a reinforcing cycle (Fig. 1), which perpetuates the effects that (indirectly) follow from these effects.

## 4.2 Structural effects

Structural effects refer to persistent changes which are observable on a macro level, including behavioural changes.

**Niche communities  $\xrightarrow{+}$  Offline social capital** The notion of social capital refers to the ability of an individual to access and draw upon the resources of the various communities that they are embedded in, especially in an offline context [39,11]. These resources can be either of a material nature, such as financial support, or of an immaterial nature, such as access to information or social support [1]. Individuals can progress online relationships from *niche communities* into the offline world to create *offline social capital*. Additionally, social networks might enable individuals to maintain access to communities after they become physically disconnected from them allowing them to benefit from maintained *offline social capital* [11].

**FoMO  $\xrightarrow{-}$  Attention span** As FoMO increases and users access social media more often, their *attention span* is diminished [29].

**Social media use  $\xrightarrow{-}$  Sleep duration** A decrease in the amount of *sleep duration* has been found to be significantly associated with *use of social media* among young adults [22]. These effects could potentially be mediated by an increased screen [17], which has purposefully been omitted from the diagram to reduce unnecessary complexity.

**Social comparison  $\xrightarrow{+}$  Mental health issues** While some research found that only upward social comparison negatively affected self-esteem [40,5], other researchers found that any type of *social comparison* on social media promoted depressive symptoms and an increase in *mental health issues* [31,4].

**Sleep duration  $\xrightarrow{-}$  Mental health issues** A lack of sleep has been demonstrated to lead to depressive symptoms and thus an overall worsening in *mental health issues* [19,38].

**Offline social capital  $\xrightarrow{-}$  Mental health issues** The accumulation of social capital has been linked to general psychological well-being [39].

## 5 Ethical responsibilities

To determine the ethical responsibilities, we examine recommendation algorithms on social media through the lens of Bose’s ethical framework of normative

business ethics theories [3]. This framework facilitates ethical decision-making for businesses by applying stockholder, stakeholder and social contract theory to produce a set obligation that businesses ought to follow to operate ethically within said theory. An overview of the ethical responsibilities per theory is provided in Table 1.

### 5.1 Stockholder theory

Stockholder theory holds that companies are responsible for maximising long-term shareholder value, while operating within the law and without engaging in fraud or deception [14]. Using the architecture decision map in Fig 1, we examine each effect to determine whether and to what degree they impact long-term shareholder value. All effects are mentioned in *italics* and can be traced back to full architecture decision map in Appendix A (Fig. 2).

**Niche communities** The offline social capital that is created through *niche communities* serves as a positive form of marketing for social media platforms, by which it improves *platform brand image* and thus increase shareholder value. Whether this effect is noticeable enough to impact shareholder value is uncertain.

**Social comparison** On one hand, social comparison is a driving factor for *FoMO*, which stimulates *use of social media*, as highlighted by the reinforcing cycle in the architecture decision map. This increases the amount of *targeted advertisements* that users are served and thus increases shareholder value. On the other hand, public knowledge of the negative effects that are result of social comparison could negatively influence the *platform brand image* and thus reduce shareholder value.

**FoMO** On one hand, *FoMO* stimulates users to make more *use of social media*, as highlighted by the reinforcing cycle in the architecture decision map. More frequent use of social media allows more *targeted advertisements* to be served to users, which increases shareholder value accordingly. On the other hand, the association of the platform with *FoMO* could negatively affect shareholder value through harming the *platform brand image*.

**Use of social media** Use of social media affects shareholder value. As users access social media more frequently, more *targeted advertisements* can be served to users, which increases *platform revenue* and thus shareholder value.

**Offline social capital** The potential increase in *offline social capital* that recommendation algorithms facilitate could positively impact long-term shareholder value by improving *platform brand image* as a marketing strategy to spotlight a positive effect of the platform. It is uncertain whether this effect is noticeable enough to impact shareholder value.

**Mental health issues** If users are critical of the *mental health issues*, this could negatively impact long-term shareholder value through damage to the *platform brand image*. The magnitude of damage to *platform brand image* is detrimental to deciding whether the deteriorating *mental health issues* of its users is the ethical responsibility of the company.

**Attention span** Similarly, the diminished *attention span* could impact long-term shareholder value, if users were to become aware of the negative effects which would damage the *platform brand image*. Again, the magnitude of damage to *platform brand image* is detrimental to deciding whether the reduced *attention span* of its users are the ethical responsibility of the company.

**Sleep duration** Similarly to *mental health issues* and *attention span*, *sleep duration* could impact long-term shareholder value, if users were to become aware of the negative effects which would damage the *platform brand image*. Again, the magnitude of damage to *platform brand image* is detrimental to deciding whether the reduced *sleep duration* of its users are the ethical responsibility of the company.

**Limitations of the theory** An important aspect of stockholder theory, is that the company operates within the laws and regulations. As this is an emerging field, there is not much regulation as of now and thus a lot is permissible within this normative theory of business ethics. This might change in the near future, as public awareness of recommendation algorithms and their effects is at an ever-increasing all time high and countries (or supranational political entities) seek to regulate them. Additionally, the magnitude of impact on shareholder value is difficult to determine for some of the effects, which complicates the process of establishing the ethical responsibilities of the company.

Ultimately, taking into account all the effects and limitations above, the deployment of recommendation algorithms on social media seems to be ethically responsible according to stockholder theory, as it increases net shareholder value, while staying within current laws and regulation.

## 5.2 Stakeholder theory

Stakeholder theory goes further than stockholder theory by stating that a company owes fiduciary responsibility not only to its stockholders, but also to anyone directly affected by its actions. Such stakeholders would primarily include: the customers, the employees, the suppliers, the wider community, creditors and the stockholders themselves [30]. Company executives and managers must ensure that company policy and business decisions not only ensure the survival of the company, but also take care of the long-term interests of each of the stakeholders [13]. In addition to the stock holders, we identified two additional groups of



stakeholders affected by the use of recommendation algorithms: the consumers and the employees.

**Social media consumers** We define social media consumers as a group of end users who predominantly consume social media content, rather than actively contributing to it. We identified several benefits as well as downsides of using recommendation algorithms to provide *tailored content* to social media consumers.

Recommendation algorithms allow users to discover and connect with other like minded individuals. Research shows that people are more likely to associate with, socialise and cooperate with those that they feel are most similar to them [6]. *Tailored content* ensures that the consumer sees content in accordance with their predicted values and personal interests. Recommendation algorithms make it easier to access *niche communities* that might otherwise suffer from a lack of discover-ability due to their small size. All of this makes it more likely that a social media consumer will actively engage with the content they see within the context of a wider community, which is one of the main goals of recommendation systems. By socialising with others in an online community the user might build up their social capital within that community. This online social capital can be potentially transformed into *offline social capital* [11].

However, some argue that the use of recommendation algorithms in social media does the opposite of promoting social interaction between its users. The use of *tailored content* runs the risk of creating so-called "*echo chambers*" [33]. These social media *echo chambers* prevent those stuck within them from being exposed to opposing view points [24]. The phenomenon of *echo chambers* might be a driving factor in political radicalisation and extremist thought [43]. By their design social media algorithms are designed to serve content that they predict the consumer will be likely to engage with. This potentially carries the risk of leading social media users down rabbit holes of increasingly more radical or polarising content [24,43].

Recommendation algorithms increase the entertainment factor of consuming content on these sites. Using *tailored content* makes it more likely that the end user finds the content more relevant and entertaining based on their own particular interests and personality traits. However, this feature of recommendation algorithms carries the potential for abuse. Mainly by facilitating the development of social media addiction [2]. This addictive use of social media use have been found to correlate with a number of *mental health issues*. [18].

**Employees** The type of employees of social media companies that we are most interested in are those actively involved in the development and maintenance of recommendation algorithms and those responsible for the development of strategies on how to implement them.

Being directly involved in the development of recommendation algorithms within a social media company carries certain reputational risks with it. In recent times the use of recommendation algorithms in social media has begun to garner

some backlash in society due to the potential negative effects it has on its users and the wider society. The reputation damage to social media companies has the potential to carry over to its employees, hurting their *offline social capital*. This is especially true for those employees who are seen as responsible for the creation and maintenance of these algorithms. They provide an easy, personified target for blame over the negative effects of these algorithms over an inaccessible social media company.

**Limitations of the theory** One of the main limitations of stakeholder theory is that it does not provide any guidance on how to rank the competing interests of the various stakeholders. This is especially problematic when these interests happen to clash with each other. There exist certain situations where the interests of a group will be affected, but it would not be prudent to include the group in the decision-making process. This is why some argue that it should be the company executives who have a final say in weighing the various competing interests of the stakeholders. However, without any reliable method or algorithm to do so, critics argue that stakeholder theory lacks any means to actually ensure the fair treatment of all stakeholders.

### 5.3 Social contract theory

Social contract theory states that enterprises should aim to promote societal welfare, by satisfying the needs of consumers and members of the society as a whole [3,7]. This hypothetical contract often has two aspects, a “social” aspect and a “justice” aspect. In short, the social aspect theorises that people will support a corporation if they stand to benefit from it. In other words, under social contract theory, a corporation should make sure the well-being of members of a society is improved. The justice aspect states that a corporation should operate in a way that avoids deception, treats members of society with respect and prevents their practices from actively worsening a group’s position within society [7].

A corporation is complying with this implicit contract, if positive and negative aspects of their operations, regarding both aspects, are balanced, or tipped towards positive. Applying social contract theory to the use of recommendation algorithms requires us to examine a lot of different values and discuss their positive and negative influence on social welfare and justice. Again, the focus will mainly be on the values associated with the use of recommendation algorithms for the use of tailored content.

**Social welfare** The main positive effect on social welfare regarding the use of tailored content, comes in the form of social interaction and social capital. We established that tailoring content can lead to the growth of niche communities and increase social interaction with like-minded people. Eventually, this can potentially translate to offline social capital as well, which is beneficial to societal welfare [39].

As tailored content drives user engagement, it also boosts the amount of time users spend on the digital platforms [15]. This increased device usage can lead to a multitude of physical and mental problems. As mentioned previously, studies have shown that *use of social media* is commonly associated with an increase in anxiety and depression among others, constituting a worsening in *mental health issues* [19,25]. A more detailed description of all the negative effects on mental health on a global scale can be found in the *Effects* section. According to the social contract theory, the companies wielding the recommendation algorithms that lead to these mental effects hold direct responsibility to these. The same kind of responsibility applies to the deterioration of healthy sleeping habits and the inducement of distorted behavioural patterns. All these effects impact the well-being of members on a societal scale.

**Justice** The most logical conclusion, when we evaluate the justice aspect, is that the use of recommendation algorithms is not ethical. Even if we assume that the corporation does its best to avoid any fraudulent and deceptive behaviour, there are still a lot of factors to argue against. Recommendation algorithms often contain implicit biases regarding gender and race [21]. This can lead to stereotyping and, for example, showing specific advertisements more often to specific groups. Depending on the algorithm’s implementation, this can reinforce existing social divisions. Recommendation algorithms also play a part in how information is distributed, often tailoring content to the particular viewpoints and personalities of the consumer. Combined with the fact that false information also tends to spread more easily on certain platforms, recommendation algorithms play a role in the increased polarisation in society, further amplifying social division in society [41]. If corporations can reliably mitigate the spread of misinformation and eliminate bias from their algorithms, it could be argued that the use of recommendation algorithms is beneficial and just. However, until corporations manage this effectively, the ethical concerns surrounding recommendation algorithms remain significant. Thus, in their current form, the use of recommendation algorithms cannot be deemed fully ethical or just from a justice standpoint.

**Limitations of the theory** In normal circumstances, a contract does not become a real contract and agreement, unless all parties agree with it. In practice, there is often a disconnect between the responsibilities members of a society put on corporations and the responsibilities corporations put on themselves. Knowing this, it is unclear how much strength can be derived from this contract, or to what degree it can even be seen as a contract. Following this, corporations might be surprised to learn that in certain situations, they are expected to make decisions in favour of societal well-being, rather than their own bottom line.

Proponents of social contract theory counter this point by stating that the strength of the theory, lies in the conceptual and hypothetical nature. Now, a social contract functions to expose the moral foundation of corporations [8]. They argue that reasoning along social contract theory would lose this main strength, it would become anything else than a theoretical contract.

#### 5.4 Responsibilities overview

Combining the effects that our within our scope from Section 1 with the results of our analysis of ethical responsibilities, Table 1 has been devised to provide an overview of the responsibilities per normative business ethics theory. Effects marked with a check (✓) are considered to be not within the ethical responsibility of the company, whereas effects marked with a cross (✗) are not. Effects marked with a tilde (~) are potentially within the responsibility of the company.

Effects	Stockholder theory	Stakeholder theory	Social contract theory
<b>Enabling</b>			
<i>Niche communities</i>	~	✓	~
<i>Social comparison</i>	✓	✓	✓
<i>FoMO</i>	✓	✓	✓
<i>Use of social media</i>	✓	✓	✓
<b>Structural</b>			
<i>Offline social capital</i>	~	✓	✗
<i>Mental health issues</i>	~	✓	✓
<i>Attention span</i>	~	✓	✓
<i>Sleep duration</i>	~	✓	✓

**Table 1.** Overview of the ethical responsibilities per personal effect as defined by the scope for every normative theory of business ethics

## 6 Regulation possibilities

**Existing regulation** In December 2023, the European Parliament publicly submitted an urging request to the European Commission to address existing legal gaps regarding digital user protection. They requested for new legislation to be introduced against the addictive design of social media. The European Commission is currently evaluating the need to update certain consumer protection legislation to ensure a high level of protection in the digital environment [28]. The European Union would be the first regulating instance in the world to tackle the addictive design of online platforms.

**Potential regulation** There are a number of different ways recommendation algorithms could be regulated to potentially decrease their addictive nature. In this section, different suggestions are described. An overview of the potential regulations and effects on the elements of the architecture decision map in Fig. 2 can be found in Appendix B (section B).

#### *Performance decrease*

The performance of the recommendation algorithms can be limited/regulated in multiple ways. This way, the addictive essence of the content is reduced. For one,

the number of variables or amount of personal data that a recommendation algorithm is allowed to perform new calculations/categorisations on can be limited. As a result, the suggested content might not always be as fitting. An alternative would be to regulate the number of recommendations that an algorithm is allowed to make. For instance, after one hundred content recommendations it is turned off for the day, and random content is displayed.

#### *Transparency*

Another area of potential regulation regards the transparency of the content recommendations and the possibility of opting out of recommendation algorithms. For now, almost all recommendation algorithms are black boxes, especially to their users. By issuing regulations regarding transparency, it might for instance become visible to a platform's users why certain content is displayed. This might induce awareness among users, potentially resulting in more responsible and limited content consumption. If the user dislikes the way they are profiled, they can then opt out of the algorithm and just be shown chronological/popular content.

#### *User control*

A compatible approach for regulation could regard the introduction of user control over the algorithmic tuning. By providing users with the ability to control the factors that influence their recommendations (e.g., adjust preferences or exclude certain types of content) users could actively steer the algorithm away from addictive content. However, increasing this control could also allow for users to make the algorithms even stronger, inducing the platform's addictive nature further.

#### *Content diversity or delay*

An alternative is issuing regulation which forces content social media platforms to occasionally replace algorithmic-generated content with random non-personalised or chronological content. In doing so, the continuous stream of personalised, addictive content is broken. This way, the chances increase that a user voluntarily disengages. Another way to induce this effect would be to add a short delay between post-consumption. This principle has been researched and has been proven to work, significantly reducing the time users spend on content-based social media [16].

#### *User awareness*

Another option is obliging content-based social media platforms to increase user awareness of their application usage. This can be done through prompts to remind users to take breaks or engage in healthy practices after a certain amount of time (e.g., "You've been scrolling for 30 minutes - consider taking a break!"). Other alternatives could be displaying a timer or end-of-session/daily summaries.

#### *User engagement cap*

A more forceful regulative option would be to cap the user engagement from the

software side. For instance, by limiting the time a user is allowed to spend on a platform. Although this would limit the 'user freedom' valued by both companies and customers, it could guarantee the reduction of user engagement.

**Proceedings** The ideal selection of regulation methods depends on multiple factors, including the effectiveness, the feasibility of enforcement, potential contradiction with current legislation and fairness among platforms [36]. Implementation of such regulation would require careful planning and communication. The European Union will set an important example in the coming years, which can provide guidelines and support for future regulators to build upon.

## 7 Discussion

The results of our study offer a method to devise ethical boundaries surrounding the use of recommendation algorithms by social media companies. We propose several ways for countries or other legislative bodies to regulate the use of these algorithms based on those boundaries. During our research, we have combined two distinct ethical reasoning models into a larger integrated project. While answering RQ1 we have identified a substantial number of effects of recommendation algorithms on the users of social media, both on an individual and societal level. We then used an architecture design decision map to model the effects of recommendation algorithms on individuals and society across several levels of deployment. Further, using the model we have shown that certain effects have an enabling or suppressing relationship with each other. We then limited the scope of our research to the effects of using recommendation algorithms to generate personalized and potentially addictive social media content. To answer RQ2 we used the three normative ethics models found in Bose [3]. We analysed the ethical responsibility of companies for using recommendation algorithms from three perspectives: stockholder theory, stakeholder theory, and social contract theory. From the perspective of stockholder theory, we found that social media companies carried very little responsibility for the effects of using these algorithms. From the perspective of stakeholder theory, we found that social media companies carried significantly more responsibility for the various effects of algorithms on the end-users of their platforms. Finally, from the perspective of social contract theory we found that social media companies carried the most responsibility as part of a larger duty to promote the welfare of society. To answer RQ3 we looked at existing and proposed regulations from the EU concerning the use of addictive recommendation algorithms. Finding that there currently is no regulation aimed specifically at addictive social media algorithms, we proposed six potential avenues for regulators to reduce the harmful effects of recommendation algorithms.

## 8 Limitations and further research directions

Our literature review was conducted in a free-flowing manner, without the use of any restricting criteria. We believe that conducting a large, structured literature review with the use of pre-defined search terms and a list of excluding criteria would have allowed us to scope our research better. Additionally due to time constraints, we had to limit our analysis of the effects listed in our architecture decision map to only one area concerning the use of personalized content. We believe that future research into the ethics of deploying recommendation algorithms should focus on other problem areas such as economic or privacy-related concerns. Lastly, during our study, we applied two distinct ethical reasoning frameworks to a single problem: the architecture decision map and the ethical business framework from Bose. We believe that these two models integrate with each other very well. A possible future direction for research would be to develop a larger meta-model combining these two ethical reasoning methods into one formal method.

## 9 Conclusion

In this paper, we have presented a new ethical reasoning method by combining two existing ethical frameworks. By combining an architecture decision map with normative theories of business ethics, the reasoning method can be used to determine the effects (RQ1) and responsibilities (RQ2) of an IT-system. Companies can apply this framework to determine whether the effects of their systems are within their responsibility and construct guidelines accordingly (RQ3). In doing so, we have demonstrated that the externalities of recommendation algorithms on social media are mostly within the responsibility of the companies that apply them for various normative theories of business ethics. To address the effects, various ethical regulatory recommendations were made, which could be implemented to reduce the negative externalities of recommendation algorithms on social media.

**Acknowledgements** This study was performed for the Responsible ICT course. We thank Sergio España and Jens Gulden for the constructive feedback throughout the research.

**Disclosure of Interests** The authors have no competing interests.

## References

1. Adler, P.S., Kwon, S.W.: Social capital: Prospects for a new concept. *Academy of Management Review* **27**(1), 17–40 (2002)
2. Andreassen, C.S.: Online social network site addiction: A comprehensive review. *Current addiction reports* **2**(2), 175–184 (2015)

3. Bose, U.: An ethical framework in information systems decision making using normative theories of business ethics. *Ethics and Information Technology* **14**, 17–26 (2012)
4. Buglass, S.L., Binder, J.F., Betts, L.R., Underwood, J.D.: Motivators of online vulnerability: The impact of social network site use and fomo. *Computers in Human Behavior* **66**, 248–255 (2017)
5. Chou, H.T.G., Edge, N.: “they are happier and having better lives than i am”: The impact of using facebook on perceptions of others’ lives. *Cyberpsychology, Behavior, and Social Networking* **15**(2), 117–121 (2012)
6. Curry, O., Dunbar, R.I.: Do birds of a feather flock together? the relationship between similarity and altruism in social networks. *Human Nature* **24**, 336–347 (2013)
7. Donaldson, T.: *Corporations and Morality*. Prentice-Hall (1982)
8. Donaldson, T.: *The ethics of international business* (1989)
9. Draper, D.: *Demystifying social media algorithms*. Bipartisan Policy Center (2023)
10. Dyer, H.T.: Defining Social Media... It’s Complicated. In: Dyer, H.T. (ed.) *Designing the Social: Unpacking Social Media Design and Identity*, pp. 15–43. Springer Nature, Singapore (2020)
11. Ellison, N.B., Steinfield, C., Lampe, C.: The benefits of facebook “friends”: social capital and college students’ use of online social network sites. *Journal of computer-mediated communication* **12**(4), 1143–1168 (2007)
12. Flayelle, M., Brevers, D., King, D.L., Maurage, P., Perales, J.C., Billieux, J.: A taxonomy of technology design features that promote potentially addictive online behaviours. *Nature Reviews Psychology* **2**(3), 136–150 (2023)
13. Freeman, R.E., et al.: A stakeholder theory of the modern corporation. *Perspectives in Business Ethics* **3**(144), 38–48 (2001)
14. Friedman, M.: Capitalism and freedom. *Ethics* **74**(1) (1962)
15. Gazi, M.A., Çetin, M., Çakı, C.: The research of the level of social media addiction of university students. *International Journal of Social Sciences and Education Research* **3**(2), 549–559 (2017)
16. Grüning, D.J., Riedel, F., Lorenz-Spreen, P.: Directing smartphone use through the self-nudge app one sec. *Proceedings of the National Academy of Sciences* **120**(8), e2213114120 (2023)
17. Hale, L., Guan, S.: Screen time and sleep among school-aged children and adolescents: a systematic literature review. *Sleep medicine reviews* **21**, 50–58 (2015)
18. Hou, Y., Xiong, D., Jiang, T., Song, L., Wang, Q.: Social media addiction: Its impact, mediation, and intervention. *Cyberpsychology: Journal of psychosocial research on cyberspace* **13**(1) (2019)
19. Keles, B., McCrae, N., Grealish, A.: A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. *International journal of adolescence and youth* **25**(1), 79–93 (2020)
20. Lago, P.: Architecture design decision maps for software sustainability. In: 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS). pp. 61–64 (2019)
21. Lambrecht, A., Tucker, C.: Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads. *Management science* **65**(7), 2966–2981 (2019)
22. Levenson, J.C., Shensa, A., Sidani, J.E., Colditz, J.B., Primack, B.A.: The association between social media use and sleep disturbance among young adults. *Preventive medicine* **85**, 36–41 (2016)



23. Montag, C., Lachmann, B., Herrlich, M., Zweig, K.: Addictive features of social media/messenger platforms and freemium games against the background of psychological and economic theories. *International journal of environmental research and public health* **16**(14), 2612 (2019)
24. Pariser, E.: *The filter bubble: What the Internet is hiding from you*. Penguin UK (2011)
25. Primack, B.A., Shensa, A., Sidani, J.E., Whaite, E.O., Yi Lin, L., Rosen, D., Colditz, J.B., Radovic, A., Miller, E.: Social media use and perceived social isolation among young adults in the us. *American Journal of Preventive Medicine* **53**(1), 1–8 (2017)
26. Przybylski, A.K., Murayama, K., DeHaan, C.R., Gladwell, V.: Motivational, emotional, and behavioral correlates of fear of missing out. *Computers in human behavior* **29**(4), 1841–1848 (2013)
27. Reer, F., Tang, W.Y., Quandt, T.: Psychosocial well-being and social media engagement: The mediating roles of social comparison orientation and fear of missing out. *New Media & Society* **21**(7), 1486–1505 (2019)
28. releases, E.P.:
29. Rosen, L.D., Carrier, L.M., Cheever, N.A.: Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior* **29**(3), 948–958 (2013)
30. Smith, H.J.: *Managing privacy: Information technology and corporate America*. UNC Press Books (1994)
31. Steers, M.L.N., Wickham, R.E., Acitelli, L.K.: Seeing everyone else’s highlight reels: How facebook usage is linked to depressive symptoms. *Journal of Social and Clinical Psychology* **33**(8), 701–731 (2014)
32. Steinfield, C., Ellison, N.B., Lampe, C.: Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of applied developmental psychology* **29**(6), 434–445 (2008)
33. Sunstein, C.: *# Republic: Divided democracy in the age of social media*. Princeton university press (2018)
34. Tandon, A., Dhir, A., Talwar, S., Kaur, P., Mäntymäki, M.: Dark consequences of social media-induced fear of missing out (fomo): Social media stalking, comparisons, and fatigue. *Technological Forecasting and Social Change* **171**, 120931 (2021)
35. Tiggemann, M., Slater, A.: Netgirls: The internet, facebook, and body image concern in adolescent girls. *International Journal of Eating Disorders* **46**(6), 630–633 (2013)
36. Trubek, D.M., Trubek, L.G.: New governance & legal regulation: Complementarity, rivalry, and transformation. *Colum. J. Eur. L.* **13**, 539 (2006)
37. Tufekci, Z.: Algorithmic harms beyond facebook and google: Emergent challenges of computational agency. *Colo. Tech. LJ* **13**, 203 (2015)
38. Twenge, J.M., Cooper, A.B., Joiner, T.E., Duffy, M.E., Binau, S.G.: Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005–2017. *Journal of abnormal psychology* **128**(3), 185 (2019)
39. Viswanath, K.: Social capital and health communications. In: *Social capital and health*, pp. 29, 260. Springer (2008)
40. Vogel, E.A., Rose, J.P., Roberts, L.R., Eckles, K.: Social comparison, social media, and self-esteem. *Psychology of popular media culture* **3**(4), 206 (2014)
41. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. *science* **359**(6380), 1146–1151 (2018)

42. Wolf, M., Sims, J., Yang, H.: Social Media? What Social Media? UK Academy for Information Systems Conference Proceedings 2018 (Mar 2018)
43. Wood, M.A.: Antisocial media and algorithmic deviancy amplification: Analysing the id of facebook's technological unconscious. *Theoretical Criminology* **21**(2), 168–185 (2017)
44. Wu, D., Zhao, Y.C.: Engagement and Addiction Dilemma Among Older Adults in Short Video Applications: Socio-technical and Self-control Perspectives. In: Gao, Q., Zhou, J., Duffy, V.G., Antona, M., Stephanidis, C. (eds.) *HCI International 2023 – Late Breaking Papers*. pp. 231–245. Springer Nature Switzerland, Cham (2023)
45. Yang, Q., Gong, X.: The engagement–addiction dilemma: an empirical evaluation of mobile user interface and mobile game affordance. *Internet Research* **31**(5), 1745–1768 (Jan 2021), publisher: Emerald Publishing Limited
46. Zhang, M., Liu, Y.: A commentary of tiktok recommendation algorithms in mit technology review 2021. *Fundamental Research* **1**(6), 846–847 (2021)

