VRIJE
UNIVERSITEIT
BRUSSEL

Master thesis submitted in partial fulfillment of the requirements for the diploma Master of Science in New media and society in Europe

Young audience's beliefs and attitudes towards YouTube recommendation algorithms

Natalia REVISHVILI

0571355

Academic year 2020-2021

Promotor: Heritiana Ranaivoson

Jury: Ana POP STEFANIJA

Social Sciences & Solvay Business School

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Tables:

Figures:

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ACM – Association for Computing Machinery

API – Application Programming Interface

AVMSD – Audiovisual Media Services Directive

AX – Algorithmic Experience

DPIA – Data Protection Impact Assessment

DSM – Digital Single Market

EUACM – ACM Europe Council Policy Committee

FTC – Federal Trade Commission

GDPR – General Data Protection Regulation

HCI – Human-Computer Interaction

ICT – Information and Communications Technology

SNS – Social Network Sites

SVD – Singular Value Decomposition

USACM – ACM U.S. Public Policy Council

ABSTRACT

In today's informationally rich world, YouTube recommendation algorithms act as an efficient tool to help users navigate through the vast array of content available on the platform. These sophisticated socio-technological systems enact great social, cultural and political power over young people and their sense-making around paramount topics. Nevertheless, their characteristics and functioning are kept undisclosed to the general public. This thesis closely examines the young audience's understanding and attitude towards algorithmic content curation on YouTube. As a result of research, several influence factors users deem to be affecting video recommendations were identified, and trends related to them were analysed. Another significant aspect studied was users' opinions on algorithmic transparency and the business rationale behind algorithmic content curation. Ultimately, this research confirmed that users already possess a good level of awareness of YouTube recommendation algorithms and their data inputs. However, their understanding of the algorithm's inner workings was limited and vague, primarily based on their observations and intuition. The second part of the research was dedicated to users' attitude towards YouTube recommendation algorithms' operation. While participants have demonstrated their awareness of concerns related to algorithmic content curation on YouTube and the business rationale behind it, they have expressed a high level of trust towards the platform and its data-sharing practices, as well as satisfaction with the quality and diversity of content recommended.

ACKNOWLEDGEMENTS

Throughout the process of writing this Master's thesis, I have received a great deal of support.

First and foremost, I would like to express utmost gratitude and appreciation to my supervisor, Dr Heritiana Renaud Ranaivoson, for his insightful comments and suggestions. Your constant guidance and valuable expertise in the field sharpened my outlook on the topic and brought my thesis to a higher level.

I greatly value and acknowledge the financial support of the Flemish Ministry of Education and Training that made my studies at VUB possible in the first place.

Finally, I want to profusely thank my husband, Raphaël Esterhazy, for his unwavering support and belief in me. I am immensely grateful for your constructive criticism and positive appreciation.

This Master's thesis came about (in part) during the period in which higher education was subjected to a lockdown and protective measures to prevent the spread of the COVID-19 virus. The process of formatting, data collection, the research method and/or other scientific work the thesis involved could therefore not always be carried out in the usual manner. The reader should bear this context in mind when reading this Master's thesis, and also in the event that some conclusions are taken on board.

Author: Natalia Revishvili

Title: Young audience's beliefs and attitudes towards YouTube recommendation algorithms

Year: 2020 – 2021

Supervisor: Dr Heritiana Renaud Ranaivoson

Words: 21958

Keywords: YouTube recommendation algorithms, user beliefs, algorithmic user experience, algorithmic content curation, algorithmic transparency.

INTRODUCTION

We live in an era of informational abundance when online platforms allow users to communicate, share and exchange opinions freely. In this informationally rich environment, algorithmic recommendation systems are meant to facilitate content discovery and information retrieval (Davidson et al., 2010).

Today, algorithms act as a powerful decision-making tool that can strongly influence our day-to-day experience with information systems (Willson, 2020).

Nevertheless, despite the sizable impact on user experience online, algorithms' nature and functioning are often inaccessible to the public and kept secret from any inspection (Pasquale, 2015). That is why recommendation algorithms are often referred to as 'black boxes', increasing the demand for transparency and accountability in algorithmic decision-making to ensure user empowerment (Koene et al., 2019).

Social media platforms have grown from a network for virtual communication to a facilitator of news consumption. Seeking economic benefit, online platforms can affect our cultural presumptions on many topics and manage our interactions with others (Gillespie, 2014). Moreover, users are constantly contributing to algorithmic mechanisms' functioning by providing them with the data inputs (likes, dislikes comments, shares, etc.) that will then be used to propose an even more personalized content (Rader & Gray, 2015).

Today, the shift of paradigm from users' consumption of information towards a prosumption implies that social media platforms are not only acting as a referral source but are also publishing unique user-generated content (Welbourne & Grant, 2016).

Being the second-largest online platform, YouTube has become extremely popular over the last decade, especially among the younger audience (Anderson & Jiang, 2018). The success of YouTube increased dramatically after its acquisition by Google in 2006, which brought along its strengths in recommendation and search

algorithms. Arguably, YouTube algorithms not only anticipate the already existing preferences of users and but also introduce discovery kind of content based on these preferences (Zhou et al., 2010). This is what allows the platform to keep users in its ecosystem for an extensive period.

Already back in 2017, the reach of YouTube was so vast that if *"YouTube's user base was a country, it would be the third-largest in the world"* (Wagner, 2019, para. 3). Around 70% of videos watched on YouTube are recommended content (Solsman, 2018). This illustrates why learning how YouTube's algorithms work, whether there is any place for bias in their functioning and what are users' attitudes towards these systems is so crucial.

YouTube's recommendation algorithms are believed to be running on a hybrid filtering principle that analyses content and user data to provide recommendations (Davidson et al., 2010). By looking at YouTube's Application Programming Interface (API), we can find several metrics that are used to generation recommendations: number of views, comments, likes, dislikes, time watched, earnings from ads, playbacks (Google Developers, 2019). However, with the constantly updating and changing nature of algorithms, it is hard to tell how these mechanisms are working at a specific point in time.

While the algorithmic content selection's purpose is to provide relevant, personalized information to users, critics are concerned that it might encapsulate viewers in filter bubbles where they will only be exposed to content that corresponds to their current views and beliefs (Bozdag, 2013).

Another concern revolves around the lack of transparency recommendation algorithms provide. While the most recent publication of Covington et al. (2016) suggests that the platform's recommendation algorithms are based on two neural networks (Covington et al., 2016), there is no blueprint for calculating the outcomes of recommendation algorithms' operations or investigating whether the approach mentioned above is still implemented. The lack of a complete information about algorithmic functioning raises a question about users' awareness and empowerment online.

Given the public relevance of YouTube's recommendation algorithms, highlighted by Gillespie (2014), it is essential to study how users, especially those who grew up with social media platforms and recommendation algorithms, understand these mechanisms and what their attitude towards them is (Gillespie, 2014).

Thus, this research aims to evaluate users' level of comprehension of YouTube's algorithms. We will study young audience's sense-making of these mechanisms and find out how they usually learn about their existence. We will also observe how they feel about YouTube algorithmic curation of the content – whether they feel empowered and in control or not and how they evaluate the extent of algorithmic transparency on the platform.

Consequently, the research questions are as follows:

RQ1: How do users understand the inner workings of YouTube recommendation algorithms?

> SQ1: How do users of the platform learn about YouTube recommendation algorithms?
>
> SQ2: In users' opinion, what data inputs do YouTube algorithms use to provide recommendations?

RQ2: What is users' attitude towards algorithmic curation of content on YouTube?

> SQ1: Do users feel empowered and in control while using YouTube?
>
> SQ2: How do users feel about algorithmic transparency on YouTube?

Before moving further with the research, several concepts need to be defined. A general notion of what algorithmic recommendations are will be explained in the theoretical framework, and different types of algorithms will be presented. We will further elaborate on social media algorithms and discuss concerns often related to them: fairness, accountability, transparency, and privacy.

This section will provide valuable context to the discussion about YouTube’s recommendation systems, their functioning, and data inputs used for their operation.

In the same chapter, we will discuss key related topics such as users’ agency and algorithmic empowerment, media, digital and algorithmic literacy, as well as the issue of power asymmetry between users and algorithms. We will also look into the concepts of users’ algorithmic experiences, folk theories and algorithmic beliefs.

This empirical research will be of explorative perspective and will take on an in-depth qualitative approach to study users’ attitudes and beliefs about YouTube’s algorithmic content curation. In-depth interviews will be carried to get an insight on participants' comprehension and feeling about algorithms adopted on YouTube. Grounded data analysis will be employed to develop a theory based on results of observation.

Finally, conclusions will be drawn and an extensive discussion of the results will be presented. Limitations and possibilities for further research will be described in this section as well.

1. THEORETICAL FRAMEWORK

1.1 Algorithmic recommendation systems

The development of Web 2.0 on a global scale created an array of new possibilities, such as simplified sharing of content, rapid access to information, and instant online communication.

It created a favourable environment for many industries to thrive, but it also brought about certain obstacles. Indeed, the ever-growing amount of digital information generated a need for mechanisms that would allow users to better navigate the digital space.

Originally, this need was met by the introduction of keyword-based search engines such as Google, although it did not yet enable the creation of an online universe where everyone would find content that corresponds precisely to their needs or preferences. At the time, the Internet was still seen as the Great Equalizer, backing a wide-spread positive image of globalization (Jeanne Dugan, 1996).

1.1.1 Definition of algorithmic recommendation systems

The problem of lack of prioritization and personalization of information was solved with the introduction of recommendation systems. Algorithmic recommendations can be described as *"tools for filtering and sorting items and information"* (Asanov, 2011, p.1). Arguably, they make life easier for everyone: while users benefit from a simplified information/product search, service providers are enjoying reduced transaction costs and increased revenues (Hu & Pu, 2009; Pu & Chen, 2010, as cited in Isinkaye et al., 2015).

Recommendation systems can also be defined as *"the use of computational processes to sort, classify, and hierarchize people, places, objects, and ideas, and also the habits of thought, conduct, and expression that arise in relationship to those processes"* (Hallinan & Striphas, 2016, p.3).

While these systems' primary function is to assist users and content providers in finding what they were searching for, their operations' main outputs are recommendations or predictions (Vozalis & Margaritis, 2003).

1.1.2 Functioning of algorithmic recommendation systems

Any recommendation system's functioning consists of three subsequent phases: an information collection phase, a learning phase, as well as a recommendation phase. The first phase is oriented towards collecting data that will serve as an input for the last phase of the recommendation process. Here, both explicit (provided by users themselves) and implicit (generated through the system monitoring of user activity) types of feedback are considered (Isinkaye et al., 2015).

Consequently, the inputs that can be “fed” to the recommendation algorithms depend on the nature of the filtering algorithms deployed. While these inputs vary, they predominantly include ratings or votes, demographic data and content metadata (Vozalis & Margaritis, 2003).

During the learning phase, the algorithms are being designed by developers to learn how to properly filter the content based on users' previous online activity (Isinkaye et al., 2015).

Finally, in the last recommendation phase, the items and information are presented to users to see what created the most significant engagement rate. This last phase acts as a valuable asset in a constant feedback loop between users and algorithms (Isinkaye et al., 2015).

Today, different algorithms filtering and sorting content are constantly upgraded and re-written to find out the most efficient way to boost user engagement and satisfaction. A brief overview of recommendation algorithms discussed in this Master's thesis can be found in Fig. 1 below.

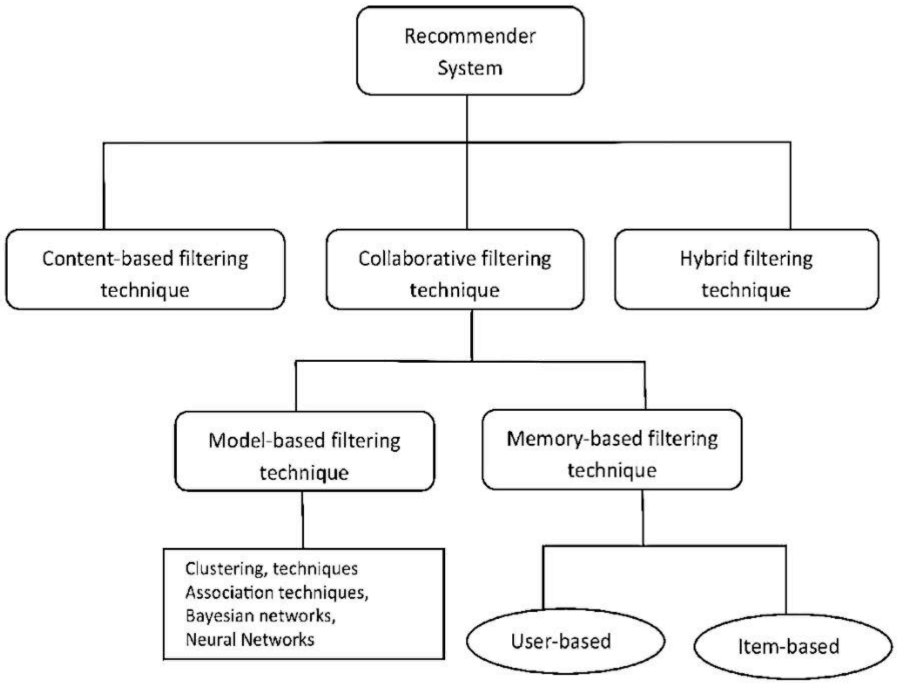

Figure 1: Algorithmic Recommendation Models (Isinkaye et al., 2015, p. 264)

1.1.3 Different types of algorithmic recommendation systems

Essentially, there are three basic approaches to the creation of every personalized recommendation algorithm that we know of: collaborative filtering, content-based filtering, and hybrid filtering (Asanov, 2011). Below, we will go through each of them and identify their strengths and shortcomings.

Collaborative filtering was first introduced in 1997 by Paul Resnick and a chief economist of Google, Hal Varian (Resnick & Varian, 1997). This algorithm relies on similar users' behavioural patterns to make predictions and propose certain content or items (Linden et al., 2003).

Collaborative filtering techniques can be divided into memory-based and model-based algorithms (Bobadilla et al., 2013). In memory-based techniques, the recommendation process is being based on ratings previously provided by users. It can either be performed through a user-based or an item-based approach. The former compares ratings of items provided by users, the latter observes similarities between these items (Isinkaye et al., 2015).

While memory-based techniques of collaborative filtering rely on data from the database directly, model-based techniques seek to build a model that will generate recommendations rather than hand-pick algorithmic inputs. If we had to compare these two collaborative filtering approaches, we would see that each of them has its strengths and shortcomings. Namely, even though memory-based algorithms are flexible and can be adapted to new data inputs, they require a long time for computing the outcomes. At the same time, model-based algorithms do not need much time to provide recommendations, but they are not as adaptive to data changes (Aditya et al., 2016).

Content-based filtering is an extension of an early version of collaborative filtering. This filtering method usually recommends items the user has already selected in the past. Therefore, there is no need to gather any data on other users to support this model (Li et al., 2012). Because almost all content-based algorithms are considering user's and item's profiles to provide further recommendations, it is possible to recommend a very niche content specific for a particular individual. It is often used when web pages, documents, or publications are to be recommended (Isinkaye et al., 2015).

One of the technique's shortcomings is its requirement of great scope of metadata on every item and a very well-structured user profile to provide relevant recommendations (Isinkaye et al., 2015). This type of algorithm also does not go beyond user preferences already indicated in their profiles. This relates to content over-specialization, a concept introduced by Zhang and Iyengar (2002) in their work on the use of linear (statistical) classifiers in model-based recommendation systems (Zhang & Iyengar, 2002).

Meanwhile, hybrid filtering combines the strengths of both content-based and collaborative filtering to compensate for their corresponding weaknesses (Isinkaye et al., 2015). The combination of methods introduces several new types of hybrid algorithmic models: "*weighted hybridization, switching hybridization, cascade hybridization, mixed hybridization, feature-combination, feature-augmentation,* *meta-level*" (Isinkaye et al., 2015, p. 263). The main reason of why hybrid filtering became so popular is its flexibility and adaptivity to systems' needs, meaning that it is possible to combine two or more recommendation methods to obtain the best results possible (Li et al., 2018).

1.1.4 Challenges faced by algorithmic recommendation systems

There are several problems commonly faced by recommender systems: the cold-start problem, the data sparsity problem, and the scalability problem.

The cold-start problem refers to a situation where a new item or user has just appeared on the platform and the amount of related data is still insufficient to generate reliable recommendations (Schein et al., 2002).

The data sparsity problem occurs when an insufficient number of items are being rated by users, meaning there is not enough reliable data to be absorbed. This jeopardizes information systems' ability to detect similarities between profiles and draw reliable connections (Chen et al., 2011).

The scalability issue pushes recommender systems to develop faster and increase their capacity to generate recommendations even when the number of datasets is limited. This is often solved with the adoption of a Singular Value Decomposition (SVD) method, in other words, the introduction of a *"low-dimensional approximation to high-dimensional data"* (Brunton & Kutz, 2019, p. 4).

While these problems are often hard to overcome using just collaborative or content-based filtering approaches, it is possible to resolve them with hybrid algorithmic models (Pandey, 2020).

1.2 Social media algorithms

In this section, we will look at the use of algorithms in media and more specifically on Social Network Sites (SNSs).

Algorithms have made their appearance into many aspects of modern society to help institutions, media channels, and companies to efficiently tackle complex challenges (as seen in Bucher, 2017; Lepri et al., 2018; Napoli, 2014; Pasquale, 2015). Today, algorithms in the media sector hold a so-called *"material agency"* over content production, distribution, and consumption, meaning that these technological tools can operate without human intervention and thus can expand and readjust on their own (Leonardi et al., 2012, p. 35).

The close link between algorithmic curation and the procession of a large and unprecedented quantity of human behavioural data makes some academicians claim that algorithms can be both cause and outcome of the Big Data or SNS phenomenon (Mayer-Schonberger & Cukier, 2013, as cited in Napoli, 2014).

The emergence of social media platforms quickly changed our lives – people became connected and separated from each other simultaneously. Algorithms are increasingly becoming a part of our everyday life and social media platforms are no exception to that (Willson, 2020). However, there are conflicting hypotheses in regards to the diversity of content promoted by recommendation algorithms.

Namely, after examining the web-browsing histories of 50.000 US citizens, Flaxman et al. (2016) came to a somewhat ambiguous conclusion. While social networks were found to promote an ideological separation among users, they also enabled more significant exposure to a greater variety of content associated with the different political, religious, or ideological groups (Flaxman et al., 2016).

At the same time, Hindman (2008) believes that traditional media forms (television, radio, newspapers) were yielding more diversity than we can find online. The author claims that search engines like Google, Yahoo! or Microsoft are narrowcasting the content we see and, by doing that, undermining our rights for democracy online (Hindman, 2008).

Social media algorithms are not only providing users with highly personalized content and filtering out unwanted content but are also helping platforms to quickly contextualize and categorize an array of data points on millions if not billions of users (Napoli, 2014).

Today, algorithms often act as a catalyst and amplifier of news creators and distributors (Napoli, 2015). Indeed, in 2012, Facebook and Twitter were among primary sources of news consumption and traffic generation for news websites (Mitchell et al., 2012).

When talking about journalism and news creation, we cannot avoid gatekeeping, which has been around ever since the 1970s’ when Dr Manning White conducted his first case study on how the news pieces were getting picked by Mr Gate (Loisen & Joye, 2017). While before, this task was solely the human editors’ responsibility, it has now mostly been entrusted to personalization algorithms that are picking and distributing the content that billions of people will see. Thus, they directly shape users' meaning-making on many crucial topics like politics, healthcare, education and more.

Moreover, with the expansion of user-generated content, SNSs have evolved from simply distributors of content to first-hand sources of information (Mitchell et al., 2012).

1.2.1 Concerns related to social media algorithms

The growing popularity of social media platforms has raised specific questions about their responsibility in front of the general public. Many of these concerns are closely related to the concept of public interest, which is often brought up when we are talking about traditional media and policymaking covering journalists’ code of conduct, media outlets’ performance, and their responsiveness to social norms (McQuail, 1992, as cited in Napoli, 2015). Some of the concerns associated with SNSs and their algorithms – fairness and accountability, transparency, and privacy – will be depicted in the following sections.

1.2.1.1 Fairness, accountability and transparency

There have always been some flaws to human decision-making, due to lack of objectivity and potential involvement of emotions. Data-driven algorithmic systems, however, were expected to avoid such distortions and allow for more objectivity and evidence-based decision-making. Yet, some claim *that "[d]iscrimination may be an artifact of the data mining process itself"* (Barocas & Selbst, 2016, p.674).

Admittedly, in 2014, the report released by the White House pointed to an evident risk of discrimination in healthcare, education, criminal justice, and employment caused by algorithmic decision-making. The authors who assembled the report offered two specific reasons for possible algorithmic discrimination: flaws in data used as inputs and shortcomings in inner-workings of algorithms themselves. As solutions to these issues, they offered several options: promotion of algorithmic fairness and accountability, a collaborative design of algorithms that would enable several stakeholders to get involved in the process, encouragement of academic research in this field, education of citizens (in other words, media literacy), and a more active role of governmental and private bodies in regulation of algorithms (Munoz et al., 2016).

Arguably, the European Union has been the most proactive in addressing serious issues related to the lack of transparency, accountability, and fairness in algorithmic systems' functioning. For instance, Article 22 of the General Data Protection Regulation (GDPR) allows citizens to demand a meaningful human explanation of automated decision-making processes enacted concerning their data. Moreover, as stated in Article 35(1), whenever data processing is likely to impose any risks to citizens' rights and freedoms, Data Protection Impact Assessment (DPIA) should be implemented (The European Parliament and the Council of the European Union, 2016).

Even though transparency might be considered a natural predisposition for algorithmic accountability and fairness and the success of any innovation, there is a wide diversity of opinions on the topic. In particular, many agree that it would be disproportionate to open up algorithms' functioning for everyone's attention.

Firstly, in a way, they are the intellectual property of online platforms and often act as considerable leverage and a part of their business model. Secondly, it is essential to ask three questions before releasing any information concerning algorithms – "how", "why" and "what". By answering each of them, it will be possible to understand the outcome of the algorithmic curation on online platforms, the algorithmic functioning rules, and access the reasoning behind the implementation of automated decision-making (Goldenfein, 2019).

At the same time, even academicians supporting the need for a more transparent operating of machine-learning systems agree that complete transparency is not possible, neither it is necessary because it will be challenging and overwhelming for laypeople to understand its coding basics anyway (Diakopoulos & Koliska, 2017).

Similarly, Zarsky (2016) admitted that while more transparency could benefit the general audience, it could also permit certain groups to access an algorithmic inner working and use this knowledge in their interest, potentially generating unwanted outcomes for minority groups (Zarsky, 2016).

Another complexity of addressing fairness, transparency, and accountability issues in regards to the algorithms is the vagueness of these terms. No one can say for sure what fairness stands for and so far, agreeing on a specific definition has proven impossible. Moreover, there is no guarantee that designing an entirely fair and a non-discriminatory algorithmic system will solve the bigger problem of discriminatory classification of users online.

Looking at the issue from a mathematical perspective, we find that taking out specific elements responsible for discriminatory practices hardly achieves any favourable results because statistically, these problematic elements will be replaced by new ones. This causes a mathematical issue because unfavourable outcomes constantly go in a loop (Goldenfein, 2019).

It became evident that purely technological tools cannot resolve the problem of the 'new digital divide', a separation between those who have access to Big Data and the ones who do not (boyd & Crawford, 2012). Therefore, the ACM U.S. Public Policy Council (USACM), in cooperation with the ACM Europe Council Policy Committee (EUACM), identified several fundamental principles of transparency, accountability, and fairness that should be considered (Garfinkel et al., 2017). These principles are as follows:

1. Awareness of citizens on algorithmic decision-making in all of the aspects of social life;

2. Access and redress that allows an investigation and correction of algorithmic systems;

3. Accountability that holds the entities implementing algorithms responsible for whatever outcomes of algorithmic decision-making;

4. Explanation that means that regardless of the complexity of algorithms, their functioning should still be possible to explain adequately and understandably;

5. Data provenance that implies transparency and accessibility of data inputs being used;

6. Auditability that guarantees an elaborate record tracking;

7. Validation and testing that ensures fairness and ethical functioning of algorithmic systems.

1.2.1.2 Privacy-related concerns

Algorithms' primary task is to save us time, assist with discovering relevant information, increase our productivity, connect us with people across the globe (Fast & Jago, 2020). However, to successfully build targeted predictive models, AI systems need a tremendous amount of personal data to absorb and process continuously, which leads to growing concerns among lay users, experts and regulators.

Tucker (2018), in her work, focused readers' attention on three types of privacy concerns frequently associated with the operations conducted by AI systems. The first one is data persistence. Once personal data is collected and processed, it is hard for users to control how long it will be stored. Many agree that due to the meagre costs of data storage, this period is extended for longer than necessary. The second concern is data repurposing, which occurs when personal data is used for other purposes than the first intended. Finally, the data spill overs concern describes a situation where people end up on databases even though they did not agree to it.

An example of that would be when someone on a photo gets recognized by facial recognition technologies (Tucker, 2018).

Another severe privacy issue is that of data breaches. An excellent example of that would be the unprecedented-in-scope Equifax data breach of 2017 when sensitive data (names, addresses, social security numbers, driver's license numbers, etc.) of millions of US, Canadian and British citizens was stolen. As was later found, the credit reporting agency knew about its security system's potential threats and weaknesses and yet did not report it or tried improving it (Leonhardt, 2019).

While it seems that big corporations cannot come back from such a scandal, the reality shows the opposite. After the well-known Cambridge Analytica incident that got tremendous momentum back in 2018, Facebook's usage was reportedly increased, and the company's stock prices got back up in a matter of only a few weeks. According to Mark Zuckerberg himself, the platform's engagement was not harmed by the Cambridge Analytica scandal as only an immaterial percentage of users deactivated their accounts as a result (Kanter, 2018).

While many believe that in the age of a highly-personalized content and algorithmic curation, "privacy is dead" (as said by Scott McNealy in 1999), Zuboff (2019) believes it is not as radical. She believes that privacy was instead redistributed, and the right to privacy is now *"concentrated within the domain of surveillance capitalism"* (Zuboff, 2019, p.15). Surveillance capitalism is a term used by Zuboff to describe how data is collected and processed by large internet giants in an undetectable and indecipherable manner to translate it into user behavior predictions necessary for targeted advertising and other purposes (VPRO Documentary, 2019).

Fast and Jago (2020) have noticed that whereas in the past political and social activist were fiercely fighting for the privacy rights of citizens, in the age of an elaborate AI functioning, the benefits brought about by algorithms dazzle people making them incapable of seeing the situation for what it is (Fast & Jago, 2020).

According to the authors, the absence of motivation to protect privacy is caused by four following factors: benefits provided by the algorithms, no expectancy of serious harm, negative consequences coming only after the service was experienced, and normalization of a lack of privacy. Authors' stance on users' attitude towards data processing is relatively negative – they doubt that people will continue prioritizing privacy and seeing it as their fundamental human right because they see the sacrifice of their data as a simple necessity in today's world. Thus, they suggest that it should be the government's responsibility to safeguard users' privacy online (Fast & Jago, 2020).

Interestingly, some noticed a shift in attitude towards privacy online – they found young people less worried about information privacy than adults. Using a 3-million-people sample to conduct a survey, Goldfarb and Tucker (2012) found that older people are less likely to reveal personal data for commercial purposes than young people. This, among other factors, can be explained by older people's tendency to be generally more private in their life (Goldfarb & Tucker, 2012).

At the same time, privacy concerns cannot be classified depending on age because they are likely to vary in different contexts. Namely, when it comes to cross-firm data-sharing and potential price discrimination, users are most likely to try and preserve their personal data regardless of age (Taylor, 2004).

1.3 YouTube recommendation systems

YouTube has shown impressive growth rates since its creation in 2005. It is by far the biggest video-sharing platform with 2 billion monthly active users and 30 million daily active users (Aslam, 2021) and with approximately 500 hours of video being uploaded on the platform per minute (Hale, 2019).

In November 2006, YouTube was purchased by Google and became one of the company's most successful subsidiaries (Erviti & Stengler, 2016). Today, YouTube is the second most-visited online platform after Google Search (Amazon Web Services, 2020). The platform is particularly appreciated by younger audiences. Notably, 85% of U.S. teens admitted to be active YouTube users (Anderson & Jiang, 2018).

The utmost importance of understanding how YouTube's algorithms work lies in fact that 70% of videos watched on YouTube are recommended content (Solsman, 2018).

The platform allows users and entities to upload, review, share, comment on music videos, talk shows, vlogs, movie trailers, and other video content types. Even though user-generated is somewhat more popular than professionally generated content, the latter is growing in volume as more and more media corporations start uploading their videos under the YouTube partnership programs (Erviti & Stengler, 2016; Kim, 2012; Welbourne & Grant, 2016).

1.3.1 Purposes of algorithmic content curation on YouTube

Like any other Big Data company (Facebook, Instagram, Netflix, etc.), YouTube's business model rests upon user engagement and an increase in the time they spend on the platform. Given that users have a limited amount of time they allow themselves to spend on YouTube, the platform developers are putting incredible effort into helping users find what they are looking for, but also to propose related personalized suggestions that are likely to keep them engaged for longer (Creator Insider, 2017).

Arguably, the recommendation systems are the best tools that help users find what they want and boost content producers' audience reach. In a way, the YouTube recommendation systems act as a gigantic feedback loop. They carefully analyse what users have already watched, what recommended videos they decided not to watch, how long it was watched to then give relevant content recommendations and keep the audience engaged for longer. For increased algorithmic recommendations' efficiency, the platform developers have decided to enable users to express their subjective opinion by pressing 'not interested' or take part in surveys where they are being asked whether recommendations were found to be relevant or not (Creator Insider, 2017). This could be considered a form of algorithmic user empowerment because users are now given a chance to largely influence their recommendations and be more active during content consumption.

On the one hand, when talking about the ethical aspect of the YouTube recommendations, many agree that even though these filtering mechanisms were accused of bias and dissemination of deceiving, conspiratory content (as seen in Alvarado et al., 2020; Lewis, 2018; Roose, 2019; Tufekci, 2018; Warzel, 2017), at the moment, it is hard to imagine alternative systems that would facilitate users' navigation in today's information overload. On the other hand, the questions about data inputs used by algorithms, transparency of their operations, and outcomes of algorithmic content recommendations are still open.

Munger and Philips (2019) in their study suggest that we should not look at YouTube as a 'hypodermic needle' which will inevitably 'inject' specific messages into people's minds but rather look at it from the supply and demand point of view. In other words, they argue that the dissemination of politically radical content is spread not only by the YouTube algorithms but also by political players who are creating these videos in the first place to influence disaffected individuals on the platform (Munger

& Phillips, 2019). Therefore, it is crucial to understand that algorithms cannot operate autonomously without any human interaction and are only a part of a much broader socio-technical system (Alvarado et al., 2020).

1.3.2 YouTube recommendation algorithms' functioning: metrics and data inputs

YouTube recommendations are not that easy to pinpoint because they are spread across different locations within the platform: home page, subscription tab, suggested videos or Watch next section, and trending videos. Interestingly, judging from what YouTube employees say, the suggested videos section located to the right of each video users have just watched is by far the most important source of recommendations (Creator Insider, 2017). Content recommended there is not just related to the video user has just watched, but also reflects his/her general interests and preferences.

According to Davidson et al. (2010), YouTube's recommendation algorithms are based on their association with a collaborative filtering principle that analyses content data and user data to select content (Davidson et al., 2010).

It is also known that throughout the period between 2005 and 2010, YouTube was solely relying on collaborative filtering, while after 2010, the platform switched to a hybrid approach which combines the strengths of both content-based and collaborative filtering to overcome potential issues. With the ever-growing amount of metadata, YouTube has also resorted to deep learning architecture that is by far the most commonly-known Machine Learning approach (Abbas et al., 2017).

It is worth noting that essentially two types of data are used by the majority of known recommendation systems: explicit and implicit. While the first type includes users' direct activity online measured in likes/dislikes, comments, and subscriptions, the second one covers factors such as, for instance, watch time and is mainly developed through the systematic monitoring of users' online behaviour patterns (Siersdorfer et al., 2010).

According to YouTube developers, the platform's algorithms rely on users' watch history, users' activity from the same digital neighbourhoods, and videos from YouTube channels they are already subscribed to (YouTube Creator Academy, n.d.).

Video metadata also plays a specific role in whether users will see a particular video in their recommendations or not. Thus, YouTubers and content creators often use keywords in their video titles, description, and/or thumbnails. This is very likely to boost their videos' reach to the target audience and to increase their ranking in the search results (YouTube Creators Academy, n.d.). After all, YouTube is now a Google subsidiary. Therefore, algorithms working astonishingly well on one platform will be very likely to be transferred to the other.

According to Alvarado et al. (2020), after reviewing YouTube's API (Google Developers, 2019, as cited in Alvarado et al., 2020), they learned that several key factors get recorded by the YouTube algorithms to generate content recommendations:

> The number of views, the percentage of viewers that the system logged in when watching the video or playlist, the number of minutes that users watched, the average length of video playbacks, and the number of comments, likes, dislikes, and shares. (p. 121:6)

Like any other recommendation system, the YouTube algorithms have to face the challenges threatening to undermine content suggestion efficiency. Namely, they have to constantly prove they can overcome the problem of scale (working with a massive corpus of videos and users), freshness (being up-to-date with everything being uploaded on the platform), and noise (external factors that might get in the way of proposing relevant videos) (Covington et al., 2016).

Typically, to successfully recommend something to platform users, the recommendation algorithms go through a two-fold process: candidate generation and a consequent ranking. All of that should be done in a matter of seconds after the search was launched by the user (Zhao et al., 2019).

1.3.3 Benefits and shortcomings of YouTube recommendation systems

This process described above cannot be done without very sophisticated mechanisms installed on the platforms to help users filter through the mass of content available online.

Over the last few years, many questions and doubts have been raised regarding the YouTube recommendation systems' bias. While this algorithmic content selection aims to provide relevant, personalized information to users, critics are concerned that this might encapsulate viewers in filter bubbles where they will only be exposed to content that corresponds to their existing views and beliefs (Bozdag, 2013). Some even argue, that YouTube algorithms contribute to the dissemination of conspiracy theories and other types of questionable content (Basch et al., 2015; Tufekci, 2018).

Moreover, social media platforms have often been found in the middle of privacy-related scandals, and YouTube is no exception to that. In 2019, the platform was accused on the grounds of unlawfully collecting kids' and underaged teenagers' data without their parents' consent. This earned YouTube millions of dollars on targeted ads. However, for the violation of data protection rules, the online giant ended up having to pay $136-million-worth of fines to the Federal Trade Commission (FTC) and $34 million to the New York city (Federal Trade Commission, 2019).

But at the same time, it was proven that thanks to the YouTube algorithms, users can enjoy a great diversity of content on YouTube through driving views to videos the algorithms believe an individual might enjoy (Zhou et al., 2010). After conducting a measurement study on a large data set, Zhou et al. (2010) concluded that despite accusations related to filter bubbles and echo-chambers, current YouTube recommendation algorithms assist viewers in the discovery of novel content that might be of interest to them (Zhou et al., 2010).

They also found that the suggested videos section is a decisive source of video's popularity and is responsible for approximately 30% of the overall number of views. A direct correlation was found between a video's view count and that of the video it was referred from (Zhou et al., 2010).

This study supports the claim that the YouTube recommendation systems are not only offering content that falls within the already-existent interests of users, but they also try to expand users' spectrum of interests and educate them.

1.4 Users' agency and users' algorithmic empowerment

The question of power is commonly addressed in social and communication studies. It is possible to differentiate between the 'power to' and the 'power over'. While the first concept refers mainly to a cooperative understanding of power, the second is more associated with the power that acts towards certain social groups' disadvantage (Loisen and Joye, 2017).

Arguably, we can trace direct or indirect exercise of power in any social or even socio-technological exchange. To understand the interaction between individuals and algorithms and analyse potential power imbalances in this relationship, we will first look into the structure vs agency dichotomy. Structure in this dialectic refers to the situation when external factors shape human behaviour. Meanwhile, agency describes the situation when individuals act under the principles of choice, free will, and individuality (Loisen and Joye, 2017).

Ideally, users are supposed to be active agents that act within a social structure and yet have the power to change it. But considering the very opaque nature of algorithmic recommendations, we might question the extent of users' agency over their online activity. This can explain a multinational demand for transparency and accountability in algorithmic decision-making to ensure user empowerment (Koene et al., 2019).

In online space, users are often referred to as prosumers implying their dual status on social media platforms: consumer and producer of the content (Toffler, 1980). This paradigm shift not only manifested the empowerment of users but also helped social media platforms grow from a referral source to a publisher of unique user-generated content, providing them with yet another valuable leverage.

This can be supported by Fuchs (2012), who analysed the political economy of Facebook and described the process of commodification of users through extraction and use of their data (Fig. 2). He clearly explained how social media platforms are exploiting prosumers to contribute to their capital accumulation and attraction of their main customers – advertisers. Users are seen as free labour generating the biggest share of surplus value and contributing to the company's long-term growth (Fuchs, 2012).

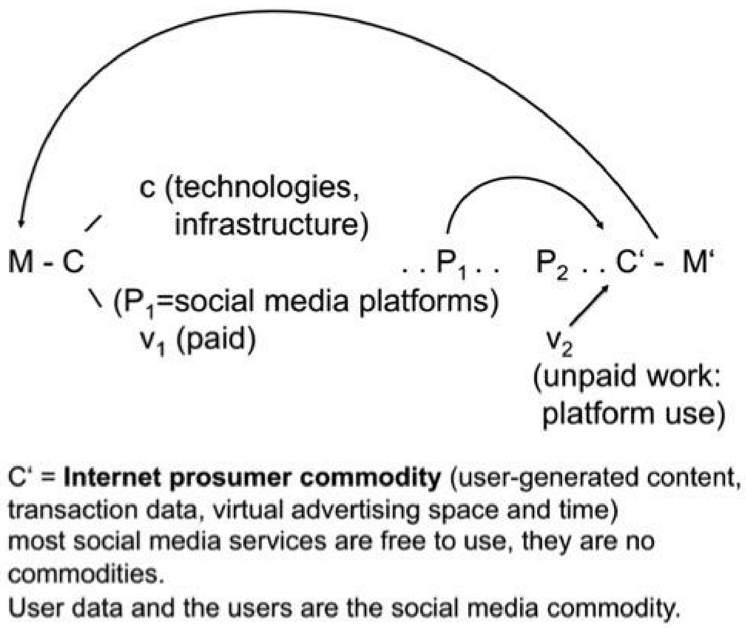

Figure 2: Capital Accumulation of Facebook (Fuchs, 2012, p. 144)

Another concept worth looking into is user empowerment. We can define empowerment as *"the process of enhancing an individual's or group's capacity to make purposive choices and to transform those choices into desired actions and outcomes"* (Alsop et al., 2006, p.1).

As mentioned above, the agency and empowerment concepts are closely linked and share multiple similarities. However, a significant difference lies in the fact that agency acts as a normative term, meaning that through gaining agency, individuals are unlikely to bring about any functional changes into their lives (Alsop et al., 2006).

Willson (2020), in her paper, emphasized the idea that algorithms are increasingly becoming a part of our everyday life, and Beer (2017) examined how algorithms are replacing human beings, asserting a tremendous social power over users (Beer, 2017; Willson, 2020). Such arguments can make us wonder whether algorithms are gaining more agency than users in the digital space.

1.4.1 Power asymmetries in a relationship between users and algorithms

With the emergence of the Internet, many were expecting it to solve existing problems and allow users to transform from passive consumers into social activists. In its early days, the Internet was seen as a tool that would allow creating an entirely new sort of public sphere that would operate outside of the outdated hierarchical system. Today, however, the power imbalances have not disappeared but simply shifted *"from the production to the filtering"* of information through algorithmic curation (Hindman, 2008, p.13).

Moreover, the very opaque and often 'black boxed' nature of algorithms diminishes users' ability to evaluate the power of algorithmic curation and shape their opinion about it (Pasquale, 2015).

The tendency towards a one-way dataveillance and extensive information gathering practices are not new and can be traced back to the 18th century when a Panopticon was first introduced (Jespersen et al., 2007). At first, the Panopticon referred to *"the all-seen"* architectural structure of prisons, which allowed the guards to watch over the prisoners without them ever knowing if they are being observed at any specific moment (Jespersen et al., 2007, p.110).

Jeremy Bentham, one of the founding fathers of utilitarianism, argued that such architectural design would allow making prisons a more humane place where prisoners would experience a feeling of constant surveillance rather than suffer from physical torture. Michel Foucault, however, broadened the concept and came up with the idea of Panopticism (Simon, 2005). This concept can accurately depict the society we live in, a society *"founded on discipline and surveillance"* (Jespersen et al., 2007, p.113).

Indeed, the key points that comprise the idea of Panopticism (Jespersen et al., 2007) expound the relationship between users and algorithms:

1. The observer is not visible from the position of the observed

2. The observed subject is kept conscious of being visible

3. Surveillance is made simple and straightforward; surveillance functions can be automated;

4. Surveillance is depersonalized because the observer's identity is unimportant

5. Panoptic surveillance can be very useful for research on human behaviour since it due to its practice of observing people allows systematic collection of data on human life. (p. 113)

It is apparent that algorithms have great power and can affect how we consume information, communicate with each other and express our opinions. In his works, Gillespie (2014, 2016) pointed out that by helping users find information, algorithms play an important role in shaping their understanding of politics and cultural life (Gillespie, 2014; Gillespie, 2016).

A similar conclusion was drawn by Willson (2020), who designated the growing power of algorithms caused by entrusting human actions to machine learning. The author apprised that a transfer of our habitus into online space is possible due to algorithms' ability to *"sort, manipulate, analyze, and predict"* (Willson, 2020, p. 3). She argues that the recommendation systems can now autonomously define how users will interact with information systems (Willson, 2020).

The role of algorithms in power dynamics was also examined by Beer (2017), who drew attention to the need of learning how algorithms work to weigh the influence they might have on users. However, as the author mentioned, due to the very complicated technical aspect of the question, it is hard for even social scientists (not to mention lay users) to avoid *"blockages in understanding"* (Beer, 2017, p.3). Thus, it is evident that to evaluate the algorithmic deployment of power, social scientists and coders need to address this question cooperatively (Beer, 2017).

It is now possible to conclude that uncertainty in how algorithms operate deprives users of their ability to exercise control over algorithmic curation. Even though the issue of algorithms having a great impact on our lives (Introna, 2011) and yet being so non-transparent has gained a certain momentum in social and Human-Computer

Interaction (HCI) studies, there is still a long way to go until all the pending questions will be answered and resolved (Beer, 2017; Introna, 2011).

1.4.2 Media and digital literacy

During the Enlightenment times, literacy referred to the ability to read, write, speak and listen. It was seen as something either empowering and freeing of limits, or as something divisive that sets different social classes apart (Livingstone, 2003).

Today, in the era of the Internet, media and digital literacy refers to the ability to understand the basic functioning of new algorithmic media and having at least some technical skills. Without this knowledge, it is nearly impossible to enjoy all the benefits of using information and communication technologies or even communicate a message in a meaningful way.

During the National Leadership Conference on Media Literacy, the concept of media literacy was defined as *"the ability to access, analyze, evaluate and create media in a variety of forms"* (Aufderheide, 1993, p. 6).

It was observed that, while most of legislative attention was focused on the *"physical connection and economic ability to access"* different sources of information, the *"literacy component"* was not considered to be anyhow decisive (Aufderheide, 1993, p.6).

Today, on the contrary, the EU legislative bodies see media literacy as an essential step in the transition to a Digital Single Market (DSM). Indeed, in the revised version of the Audiovisual Media Services Directive (AVMSD) of 2007, we can find the additional Article 26, requiring the European Commission to *"submit to the European Parliament, the Council and the European Economic and Social Committee…on the levels of media literacy in all Member States"* (The European Parliament and the Council of the European Union, 2007, L332/44).

Some might say that we live in times of disruption when new technologies are dramatically changing our lives and the ways we communicate, solve problems,

work, collaborate, and consume information (Saputra & Al Siddiq, 2020). Thus, today, digital and algorithmic literacy have become necessary skills for an individual's successful socio-economic integration.

Lanham (1995) argues that the ability to write and read is not enough anymore to be considered literate. It is also crucial to have skills that will allow conceiving the message and its meaning regardless of the form in which it is presented (Lanham, 1995, as cited in Lankshear & Knobel, 2015).

Digital literacy cannot be referred to as a specific set of skills. It is rather the ability of individuals to adapt to the new media environment and use all information sources available to meet their needs (Gilster & Glister, 1997, as cited in Isaias et al., 2013).

To highlight the diverse nature of digital literacy, van Deursen and van Dijck (2009) clustered Internet or digital skills into four different categories: operational skills (the ability to operate freely in digital media space), formal skills (the ability to understand and use specific digital structures, i.e. URLs, menus), information skills (the ability to find specific information online), strategic skills (the ability to apply knowledge on practice and satisfy personal needs) (Van Deursen & Van Dijk, 2009).

Recently, there has been a rise in public interest in algorithmic biases such as filter bubbles. Even though the algorithmic content selection is supposed to assist in suggesting highly personalized content to users, some experts are convinced that algorithms are likely to encapsulate viewers in filter bubbles where they will only be exposed to content corresponding to their interests and opinions (Bozdag, 2013). This, as a result, might restrain users' access to opposite opinions, lead to a lack of information diversity and impose a direct threat to democracy (Bozdag & van den Hoven, 2015; Rader et al., 2018).

In conclusion, media, digital and algorithmic literacies act as a valuable addition to the initiatives demanding more algorithmic transparency, fairness, and accountability. Only when all of these components are working together can we expect positive outcomes regarding online social participation and user empowerment.

1.5 Users' algorithmic experience

Algorithmic experience can be understood as a framework that evaluates users' interaction and experience with the algorithms (Alvarado & Waern, 2018). It can often be implemented to understand how users' personal experience with the platform can influence their activity online. Thus, after accessing platform's interface and algorithmic experience (AX) of users on Netflix, Alvarado and Waern (2018) came up with a framework that could improve users' algorithmic experience and resolve the lack of user empowerment issue. The model included seven consecutive points: *"algorithmic profiling transparency, algorithmic management, algorithmic awareness, algorithmic user-control and selective algorithmic remembering, algorithmic usefulness and algorithmic social practices"* (Alvarado & Waern, 2018, p. 18). Each of these characteristics were believed to enable platforms to improve algorithmic profiling transparency by offering several categories of movies rather than single movies individually, allowing users to gain more control over recommended movies and being transparent in showing specific inputs algorithms are using to generate movie recommendations (Alvarado & Waern, 2018).

Furthermore, the algorithmic awareness of users can directly affect the nature of their algorithmic experience. To examine how users' perceived knowledge about Facebook's algorithms can potentially influence their behaviour on the platform, Eslami and his colleagues (2015) conducted a study of 40 lay Facebook users. According to study, we may observe a generally low level of users' awareness of algorithmic curation, which does not necessarily come from the lack of users' interest in this matter but can be explained by the 'black boxed' nature of algorithms. In turn, the study revealed that user awareness of algorithmic curation can drastically change how they choose to engage with the social media platform – how they interact with the platform's settings and/or communicate with other users on the platform (Eslami et al., 2015).

Arguably, our AX can shape our perception of both online and offline realities. To illustrate it, Bishop (2018), in her work, suggested that YouTube's algorithms could cause gender-related polarization and bias. After selecting beauty vloggers as her case study, Bishop noticed that YouTube algorithms recommend very feminized content to female audiences, promoting hegemonic ideas of how females should look and act nowadays. As a result, in some instances, YouTube algorithms and

YouTube culture might be a catalyst for feelings of anxiety, insecurity, and discomfort (Bishop, 2018).

To overcome these negative consequences of algorithmic curation and propose possible solutions, it is essential to understand users' sense-making and attitude towards algorithms. This might be part of the reason why several academicians (such as Baumer, 2017; Diakopoulos, 2016; Lee et al., 2017; Oh et al., 2017) are highlighting the importance of conducting more human-cantered studies about human-algorithm interactions.

Users' perception and understanding of algorithms, among other theoretical frameworks, can be studied through the examination of folk theories and user beliefs (Alvarado et al., 2020).

1.5.1 Folk theories

French and Hancock (2017) described folk theories as an *"implicit collection of beliefs about the system"* (French & Hancock, 2017, p. 5). They compared folk theories to the often-related notion of mental models, which are mostly used in the HCI research. Even though these two concepts share a similar underlining, they still differ in their essence. While mental models are built on user's previous experience and interaction with the cyber-social system, folk theories represent the intuitive understanding of platforms' inner workings, which might then guide users' behaviour online (French & Hancock, 2017).

DeVito et al. (2018) tried to link folk theories with users' self-representation on social media platforms (DeVito et al., 2018). In their work, they aimed at understanding how the process of folk theories formation might be explained in relation to algorithmic social media recommendations. They drew two main results. Firstly, it was concluded that knowledge of how algorithms function might not only contribute to a better user experience but can also positively influence users' self-representation online. As a second result of the research, they suggested that platforms themselves incorporate folk theories to target problematic aspects of users' interaction with the platform (DeVito et al., 2018).

The concept of folk theory is often related to consciousness, intentionality, and free will. Thus, we may conclude that without having folk theories about something, an individual cannot construct his/her opinion and act upon it. According to Malle (2009), the folk theory of consciousness is based on three mental actions:

monitoring, choice, and subjective experience (Malle, 2009).

When users cannot access the functioning of recommendation algorithms on YouTube or any other modern social media platform, they, by definition, cannot behave consciously and/or act intentionally. As depicted in Fig. 3, an action can be considered intentional only if an individual has expressed a desire for it and belief that an action taken will result in a desirable outcome. Thereupon, intention can lead to the state of intentionality when it is combined with a sense of awareness and a set of skills necessary for the realization of the initial intention (Malle, 2009).

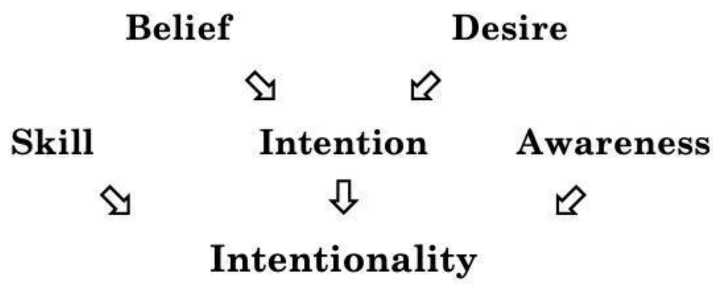

Figure 3: The Folk Concept of Intentionality (Malle, 2009, p. 4)

1.5.2 Algorithmic beliefs

User beliefs are part of a feedback loop where they act as both input and outcome of algorithmic functioning. That is why user beliefs can directly influence users' interaction with the platform and shape the functioning of the system as a whole (Rader and Gray, 2015). Notably, even without any prior knowledge on how the system operates, users can still shape their beliefs based on their personal experience with it (Gelman & Legare, 2011).

By conducting a qualitative research of users' awareness on algorithmic curation of the content, Rader and Gray (2015) developed several categories of users' beliefs about Facebook News Feed. 'Passive consumption' illustrated no awareness of recommendation algorithms being implemented. 'Producer privacy' was related to the belief that what users see is influenced by their friends on the platform. 'Consumer preferences' category depicted the belief that *"the News Feed will not show them exactly what they want to see"* (Rader & Gray, 2015, p. 6). Moving further, 'missed posts' underlined the idea that the platform is hiding certain posts without users' consent. 'Violating expectations' explained the confusion users might feel when expecting the recommendations to show certain posts but instead seeing different ones. And lastly, 'speculating about the algorithms' delineated users' belief that what they see in their News Feed is fully curated by the platform's algorithms.

The classification of user beliefs offered by Rader and Gray is relevant because it can act as a valuable ground for discussing new design possibilities that can improve user experience with the platform.

In their attempt to study user beliefs about the video recommendation systems and how they might influence users' interaction with the social media networks, Alvarado and his colleagues (2020) conducted semi-structured interviews with middle-aged YouTube regular users who do not have any prior education in Information and Communications Technology (ICT). As a result, four main categories of user beliefs were drawn: *"Previous Actions, Social Media, Recommender System, and Company Policy"* (Alvarado et al., 2020, p. 121:2). The findings of interviews suggested that even though these users knew about the existence of algorithmic curation on YouTube, there was no exhaustive understanding of how the algorithmic mechanisms function (Alvarado et al., 2020). The study's primary limitation is that they mainly interviewed users between 37 and 43 years old, which limited the generalizability of findings. The research did not include the algorithmic experience of a younger audience who grew up with social media networks and could have demonstrated different levels of knowledge. This shortcoming, however, is being addressed in this Master's thesis.

2. METHODOLOGY

In this chapter, we will examine the methodological design of the thesis, introduce interview questions and define the sample characteristics.

2.1 Research rationale

This thesis takes up an explorative perspective in an attempt to study users' attitudes and experiences, as well as an interpretive perspective because the research aims at understanding the subjective sense-making of the phenomenon by participants. In order to provide better context, we will also develop concepts such as algorithmic experience and user empowerment through critical research. However, users' beliefs and attitudes towards algorithmic curation on YouTube will remain the central focus of the thesis.

2.1.1 Aim of the research

The aim of this research is to evaluate regular users' level of comprehension of YouTube's algorithms – study their sense-making of these mechanisms, find out how they usually learn about their existence; as well as to see how they feel about YouTube algorithmic curation of the content – whether they feel empowered and in control and how they feel about algorithmic transparency on the platform.

To be more precise, with this thesis, the researcher aimed to study the beliefs and attitudes of young adults born from 1985 onward, who can, therefore, be related to Millennials' generation and/or Generation Z (Pew Research Center, 2010; Dimock, 2019). These young adults are familiar with the Internet and have grown up with social media networks or seen their rise during their adolescence, which can influence their digital behaviour and algorithmic awareness.

2.1.2 Research gap and scientific purpose

As algorithms grow deeper into all realms of human lives, it is important not to limit this field of study to only the technical aspect of algorithmic functioning. People get largely affected by recommendation algorithms on YouTube and other SNSs. And yet, it seems that not enough attention was paid to user-centred approach to an evaluation and analysis of algorithmic functioning.

At the same time, while several scholars have already covered the topic of users' beliefs about Facebook's or Tweeter's algorithms (such as Diakopoulos, 2016; Berg, 2014; Bucher, 2017; Barberá et al., 2015; Chaffey, 2016; Nunez, 2016; Rader & Gray, 2015; Tufekci, 2015), it is not yet the case for YouTube.

The expanding library of educational videos and easy navigation across the platform makes it attractive for young people to resort to YouTube as one of the main resources for gaining knowledge (Antonio & Tuffley, 2015). YouTube's expanding role in guiding users' opinions and acting as an educational tool on crucial topics is yet another reason to study YouTube recommendation systems and users' understanding of them.

Existing works about the user aspect of social media algorithms are still scarce, something that might be explained by the topic's novelty. At the same time, studying users' understanding of the YouTube algorithmic functioning has a tremendous relevance because this can largely shape their interaction with the recommendation systems and, thus, their agency over their media consumption. The fact that around 70% of videos watched on YouTube are recommended content (Solsman, 2018) can explain the importance of researching users' understanding and attitude towards algorithmic curation on the platform.

Therefore, this Master's thesis aims to contribute to the already existing knowledge on the topic by studying young audience's interpretation of YouTube algorithms, as well as to probe deeper into users' attitudes and beliefs towards YouTube algorithmic content curation, its treats and transparency.

2.1.3 Research questions

The research questions and sub-questions are as follows:

RQ1: How do users understand the inner workings of YouTube recommendation algorithms?

SQ1: How do users of the platform learn about YouTube recommendation algorithms?

SQ2: In users' opinion, what data inputs do YouTube algorithms use to provide recommendations?

RQ2: What is users' attitude towards algorithmic curation of content on YouTube?

SQ1: Do users feel empowered and in control while using YouTube?

SQ2: How do users feel about algorithmic transparency on YouTube?

The first research question and two of its sub-questions will allow us to examine young audience's understanding of YouTube algorithm functioning, gain insights on their reflection of the technical construction of YouTube algorithms by recording which data inputs they think are included in this algorithmic process and are the most/least relevant. This will allow us to look into respondents' algorithmic beliefs and folk theories.

The second question and related sub-questions are focused more on users' feelings and attitudes about recommendations on the platform. These questions, therefore, will assist in concluding on whether users feel empowered by using the platform or not, as well as get valuable insights of what is their opinion about algorithmic transparency on YouTube.

2.2 Data collection process

In this sub-chapter, we will examine the data collection method chosen for this thesis.

2.2.1 In-depth interviews

For the data collection method, the researcher decided to opt-in for in-depth qualitative interviews with regular YouTube users aged 18-35. This allowed obtaining an in-depth insight on the sense-making of the YouTube recommendation systems among participants who fitted sampling requirements and study their attitudes towards algorithmic curation of the content.

The semi-structured format of interviews was chosen because it allows covering all the important questions and yet not limit interviewees in freedom to expand the topic and express their feelings, which, in turn, might result in a gradual expansion/editing of the interview questions (Mason, 2002). The average duration of the interviews was approximately 30 minutes. All interviews were audio-recorded and transcribed.

Due to COVID-19 safety measures and gathering restrictions, it was decided that online interviews via Microsoft Teams were the most appropriate option. Real-time communication allowed discussing the topic more openly and freely and effectively applying probing strategy to expand the discussion further.

As research ethics imply, it was essential to ensure that the researcher respects participants' needs and interests and is not violating their right to privacy (Flick, 2009).

All participants were granted total anonymity throughout the coding process – only numerical codes were used instead of participants' real names. This ensured that no moral rights were violated and participants could feel safe and comfortable sharing their opinions on the topic.

A couple of days before each interview, an informed consent form was handed to participants to make sure that they agree to the procession of the responses within this research without a possibility of transfer of any of their personal information to third parties. It also allowed the researcher to make sure participants knew the purpose of the research and realized that they could stop the research process at any time if they felt uncomfortable (Bryman, 2016).

Moreover, all participants were treated equally and with respect for their opinions, which ensured that the justice of this research was preserved.

2.2.1.1 Interviewing process description

For the interviewing process, a combination of non-biased and suggestive methods of questioning were employed.

The interviews were conducted in the period between 22 of February 2021 and 11 of March 2021. All participants were asked to provide their demographic details: name, surname, gender, occupation, country of origin, and contact details in the drop-off questionnaire that was given beforehand. A sample of the questionnaire and the informed consent paper can be found in the appendix list (see: Appendix 1 and Appendix 2).

In-depth interviews were following a semi-structured set of topics: (1) awareness on the existence and the purpose of YouTube algorithms (2) perception of YouTube algorithms' functioning (3) attitudes towards algorithmic curation of the content and algorithmic transparency (for the interview questions, see: Appendix 3).

During the first few minutes of the interview, the researcher quickly reminded the topic of the discussion and introduced herself. The idea that it is a not knowledge-based interview and is instead oriented towards learning users' subjective experience was stressed. It was seen as an essential step towards establishing a trustworthy, natural ambience between the researcher and the participant.

Although projective techniques were often implemented to expand the conversation, there was still a clear structure to the interviewing process. Namely, while the first two question topics were primarily concentrated on users' awareness and subjective understanding of YouTube algorithmic curation, the third set of questions was aimed at learning users' feelings on the existence of YouTube recommendation algorithms, their transparency, as well as concerns they might have associated with algorithmic functioning on YouTube.

2.2.1.2 Sampling process and sampling characteristics

To comply with the aims of the research, judgment or purposive sampling was selected. This approach enabled the researcher to select participants that correspond to all initial criteria of the sampling.

To avoid ethical issues and obtain a socio-economic homogeneity, the sample was comprised of young adults between 18 and 35 years old. The regional aspect of the sample's demographics was not decisive in this thesis, this helped maintaining a sample diversity. The researcher made sure that the sampling process was not in any way discriminatory on gender, religion, ethnicity-basis.

As there might be a bias regarding knowledge held about the functionality of the recommendation system, all participants with an extensive academic background in ICT or computer sciences were excluded. The research was also focused on participants who use YouTube for leisure and/or educational purposes only and might be regarded as YouTube regular users (spending at least 60 minutes on the platform per week). Participants should not have earned any money by uploading content on the platform, this allowed the researcher to exclude those who perceive themselves as 'YouTubers' or video-content-creators. The socio-demographic characteristics of participants are presented below:

Participant	**Gender**	**Country of origin**	**Profession or field of study**	**Age**	**Use of YouTube**	**Comments**
P01	female	Georgia	Business manager	23	17h	Music, TV shows, online workshops and webinars
P02	female	Germany	Student (MSc): Communication Studies	26	1h – 2h	DIYs, makeup tutorials, music and music videos, fashion vlogs
P03	male	Belgium	Student (MSc): Communication Studies	21	10h	Music, video games and education

P04	male	UK	Communications assistant/client relationship	25	10h	Entertainment shows
P05	male	Belgium	Freelancer/digital marketing specialist	31	4h – 5h	Education and philosophy
P06	female	Italy	Student (MSc): Communication Studies	24	3h – 4h	Education, politics, sports
P07	female	Brazil	Student (MSc): Communication Studies	28	3h	TV shows, celebrities' interviews
P08	female	Poland	Student (MSc): Communication Studies	24	14h	Music, makeup tutorials, ASMR
P09	female	Switzerland	Student (MSc): Communication Studies	26	2h	Comedy, education, yoga and well-being
P10	female	Ukraine	Student (MSc): Communication Studies	24	5h	News, education, sports and meditation
P11	male	Germany	Student (MSc): Communication Studies	25	1h	Comedy shows
P12	female	Brazil	Student (MSc): Communication Studies	25	3h	Cooking, planting tutorials, news and politics
P13	female	Georgia	Student (MA): Project Management	23	6h	Music and entertainment
P14	female	Lithuania	Student (MSc): Communication Studies	24	1h – 1.5h	Beauty and education
P15	female	Belgium	Student (MSc): Communication Studies	22	1h	Education

P16	female	Russia	Student (MSc): Communication Studies	30	2h	Music, education and entertainment
P17	male	Belgium	Musician/music production	30	4h – 7h	DIYs, music production tutorials, entertainment and comedy show
P18	male	Scotland/ Germany	Student (MSc): Communication Studies	25	8h – 10h	Sport shows, entertainment, education and documentaries
P19	female	Italy	Student (MSc): Communication Studies	27	3h	Music and education
P20	male	Belgium	Product owner	32	7h	Music, entertainment, music production tutorials, history-related content, football

Table 1: Participants' socio-demographic characteristics

2.3 Data analysis process

As mentioned above, all interviews were audio-recorded (with the participants' consent) and then transcribed.

The researcher decided to resort to Grounded theory approach because of the inductive nature of the thesis. Introduced by Glaser and Strauss, the Grounded theory approach allows to conduct data analysis and collection simultaneously, and further develop categories, sub-categories, and, ultimately, build a theory. Three phases of coding were utilized: open, axial, selective (Strauss & Corbin, 1990).

During the open coding phase, the researcher strived to analyse the transcripts and draw categories from the data. Further, during the axial coding phase, categories and subsequent sub-categories were developed. Finally, during the selective coding process, categories and sub-categories were analysed and discussed in a way that allowed answering the research questions (Strauss & Corbin, 1990).

The developed coding structure can be found in Appendix 4.

3. DISCUSSION OF RESULTS

Discussion of the results will follow the same chronological order as the research questions:

RQ1: How do users understand the inner workings of YouTube recommendation algorithms?

> SQ1: How do users of the platform learn about YouTube recommendation algorithms?
>
> SQ2: In users' opinion, what data inputs do YouTube algorithms use to provide recommendations?

RQ2: What is users' attitude towards algorithmic curation of content on YouTube?

> SQ1: Do users feel empowered and in control while using YouTube?
>
> SQ2: How do users feel about algorithmic transparency on YouTube?

To lay the ground for further analysis, an introductory overview of users' motivations to use YouTube, as well as their perceived advantages and disadvantages of the platform will be presented.

3.1 Purposes of using YouTube, advantages and disadvantages

There were three main purposes participants claimed to be using YouTube for: leisure and entertainment, news retrieval, and education (which included both academic and life-applied purposes). The weight of each of the categories is illustrated in Fig. 4 below:

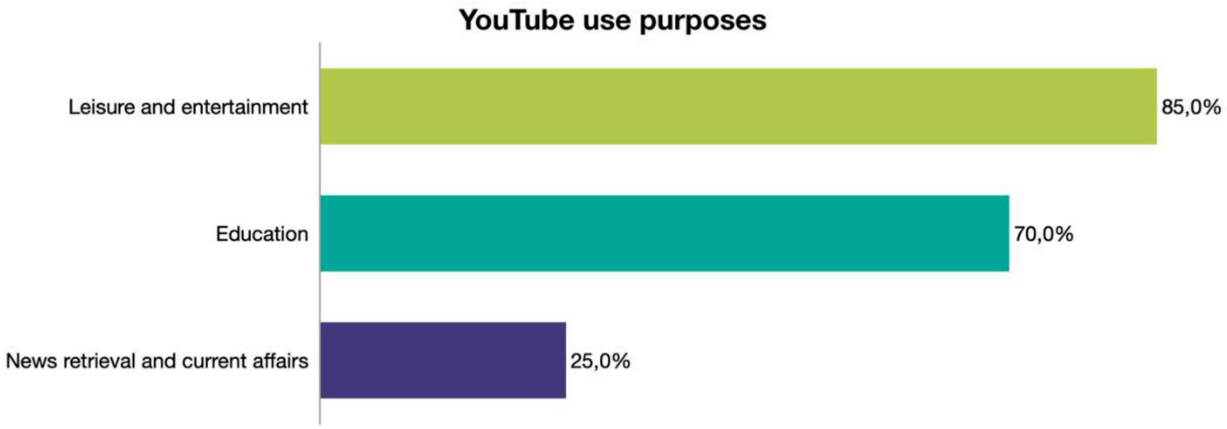

Figure 4: YouTube use purposes

Among many leisure purposes (stand-up comedy videos, TV shows, interviews, movie trailers and overviews), music had by far the largest share. Even after an introduction of alternative music platforms, YouTube remains the place where users expect to find a great variety of music for free:

"Mostly I use it [YouTube] for music. I don't have Spotify or anything like that. I guess it is because of, you know, it's for free and really don't feel like spending money on music" (P08, 24, female, MSc student, Polish).

However, YouTube appeared to be more than a source of entertainment and a way to spend free time. As many of the participants were away from home, they resorted to the platform when they wanted to get acquainted with the latest political news or current affairs from their home country. Some participants saw the platform as a valuable and, most importantly, trust-worthy source of information:

"The first thing I look for there [YouTube] is news. For example, I like short explanatory videos from the channels I am already subscribed to because I usually know I can trust the author of the video and I can expect a high-quality, fact-checked content" (P10, 24, female, MSc student, Ukrainian).

Lastly, participants were often looking for video content that would allow them to learn a new skill, understand a new topic, deepen their knowledge of a specific field.

Some students also claimed to be accompanying their studies with educational YouTube videos:

"There's a lot of great content coming from independent creators that offer a variety of high-quality explanatory content. And often before looking for information in official papers, I first go to YouTube, not even Wikipedia" (P09, 26, female, MSc student, Swiss).

When asked about the advantages of the platform, the majority of participants pointed out to a variety of content that might contribute to their knowledge or assist in spending their leisure time. Whenever using the platform, respondents were assured they would find precisely what they were looking for. During the suggestive part of the interview, 60% of respondents agreed that YouTube algorithms were not only suggesting videos similar to what they had already watched but also offered slightly novel content:

"I think that they are trying to propose something new, but I think they still try to stay in a safe zone as much as possible. Therefore, even if they would choose something new and not directly related to what you have already been watching, that will still be something that they think you could like, or you could click on" (P19, 27, female, MSc student, Italian).

This supports the findings drawn by Zhou et al. (2010) in claiming that YouTube recommendation systems are not only guiding users in a familiar direction but are also increasing video diversity on the platform by suggesting new discovery type of content to users (Zhou et al., 2010).

Some respondents pointed out their own time-wasting behaviour, which they felt was enhanced by YouTube recommendation algorithms. They expressed their concerns about the risk of spending too much time on the platform without any particular reason. Moreover, as the interviews were conducted during the COVID-19 crisis, participants declared that they wanted to avoid spending more time watching the screens:

"Now it just feels like I want to do other things with my time than watch YouTube. Also, now, during pandemics and a lockdown, we constantly watch our screens and then spending another hour doing that while watching videos annoys me in a way. There is also a lot of useless content on there and I don't want to waste my time, I'm trying to make a conscious decision" (P04, 25, male, client relationship and communication assistant, British).

3.1.1 Engagement and (dis)satisfaction with the recommended content on YouTube

To better comprehend users' grasp of YouTube algorithms, participants were asked to describe their engagement with the recommended content on YouTube and evaluate its relevancy.

The interviews' results revealed that users are much more likely (60% of respondents) to click on the videos recommended in the suggested videos section located to the right from the video just watched, than the ones on the home page. This can be explained by targeted use of YouTube that consists of searching for a specific topic and then expanding from it:

"I just type in the search bar what I am looking for and that is it. Therefore, I don't even pay much of attention to my homepage to be honest. But then the suggested videos I often find to be quite relevant so I might often click on them" (P07, 28, female, MSc student, Brazilian).

Another explanation to that can be an attempt to save time and not fall into an endless span of videos recommended on the home page:

"I would mostly click on the suggested videos rather than those that are presented on the home page because I don't have all the time in the world to spend watching YouTube videos" (P05, 31, male, digital marketing specialist, Belgian).

This observation supports findings of Zhou et al. (2010) who claimed that suggested videos section located to the right from the video just watched is the primary booster of video’s views on YouTube (Zhou et el., 2010). The same conclusion was also introduced by a Product Manager of YouTube search and discovery team, Todd Beaupre (Creator Insider, 2017).

Meanwhile, when it comes to a more entertainment-oriented content, users were found to be more likely to resort to the home page suggestions because they knew it is going to be perfectly-suited for their needs and will save them time for searching:

“…I am rarely looking for a very specific content, I am going on YouTube to rather watch something fun or watch a TV show like the Voice. Therefore, when I am on YouTube I already know that there is nothing I need to search, lots of videos will already be there on my home page” (P11, 25, male, MSc student, German).

We could conclude that participants mostly enjoyed the quality and diversity of content that gets recommended to them by YouTube. Namely, in one of the interviews, YouTube was referred to as *“the biggest library of video information”* (P05, 31, male, digital marketing specialist, Belgian).

3.2 Awareness and personal experience with YouTube recommendation systems/algorithms

Before moving any further, it was essential to first probe into users’ awareness of YouTube recommendation algorithms' existence. Here, participants were asked to reflect on how they think certain videos appear in their YouTube recommendations feed, whether that is a random process or guided by specific mechanisms. This part of the interview was very broad and non-suggestive which allowed the researcher to understand where each participant stood in terms of his/her reasoning behind the YouTube algorithmic functioning.

3.2.1 Users’ awareness on YouTube algorithms

All 20 participants showed absolute confidence in stating that content on YouTube is curated and is not a result of a random video selection. However, the level of understanding and the way of learning about YouTube algorithms differed among them.

While most users displayed a high level of awareness, their conception of algorithms was very intuitive and sometimes poorly specified. Namely, only half of participants referred to the term "the algorithms" or "YouTube algorithms":

"Of course, I don't think that it is a random process. I can personally notice that videos I see on my home page or in recommended videos section are very much related to the videos I've seen in the past, past few hours or a few days…The algorithms have learned and analysed my preferences to then provide me with a very personalized selection of videos, that is how I see it" (P13, 23, female, project manager, Georgian).

Another half used the term "the mechanism" mentioned by the researcher. The terms such as "collaborative filtering", "content-based filtering", "hybrid filtering" or "machine learning" have not appeared in any of the interviews.

The findings listed above, in a way, resemble with some of the earlier research on algorithmic awareness on YouTube by Alvarado et al. (2020), where they aimed at studying middle-aged users' perception of YouTube algorithms. As the results of this study showed, respondents either detected some level of content curation on the platform, or learned about it from media sources. However, they were lacking a clear vision of how exactly YouTube algorithms work and what data is used to keep them running. Arguably, this can be linked to users' general lack of interest in how these systems operate or to the complex technical nature of the process.

Meanwhile, in this study, despite a lack of in-depth technical understanding of algorithmic functioning, users showed a well-rounded understanding of how one's engagement with the platform can influence future content recommendations and presented an extensive knowledge on which data inputs can be analysed by YouTube algorithms. Our users have also shown an understanding of how it is possible to manage the content they get recommended and change recommendations if necessary. This could be, in turn, explained by the fact that respondents grew up with SNSs and already had at least some understanding of algorithmic functioning. Moreover, an increased media interest in the topic, as well as recurring scandals related to algorithmic biases could act as a catalyst for some of the interviewees' curiosity.

3.2.2 Users' means of influencing their YouTube recommendations

Almost all participants displayed a high level of awareness and personal use of mechanisms that might allow them to gain control over recommended content. Three categories of such mechanisms were identified by the end of the interviewing process.

Firstly, most respondents mentioned an option of changing search queries and transforming their usual online behaviour patterns. Following this path, they believed, it is possible to gradually steer algorithms in a new direction:

"If I would put a lot of work into it, I might be able to change it depending on what I search and what I start looking for. Then after the algorithms will for sure adapt to this new type of behaviour" (P02, 26, female, MSc student, German).

An alternative option was introduced by two of the respondents – they offered to erase or deactivate watch history analysis altogether. This, they proclaimed, will make algorithms less adapted to one's preferences and interests but, at least, will make users more in control of what they see on the platform:

"I just deactivated the function on my watch history analysis. That shows that as a user you still have some power over your actions on the platform in your hands" (P15, 22, female, MSc student, Belgian).

Finally, it was reported that editing one's preferences or interest categories in Google account linked to YouTube account can act as another recommendation-managing tool:

"I know that I can always change my interests and preference on my Google account. It's not very promoted or not communicated enough but you have lots of control over your privacy on your Google account, as well over your interests, what types of ads are appropriate for you and which are not, what do you want to be recommended and what not" (P05, 31, male, digital marketing specialist, Belgian).

Another respondent not knowing whether such option exists on YouTube, drew a parallel with a different social media platform, Facebook, and assumed that it works the same way:

"I just don't know how it technically works on YouTube, but on Facebook you can even go into your interest categories and change them, even delete them if you want" (P11, 25, male, MSc student, German).

Users' advanced understanding and exploitation of mechanisms that allow them to manage their recommendations could signal their high level of awareness on YouTube algorithmic functioning, according to Eslami and his colleagues (Eslami et al., 2015). This theory could be linked to user empowerment. While it consists of users' ability to transform their deliberate choices into actions (Alsop et al., 2006), it cannot exist without users understanding the systems' inner workings. This question, however, is extensively addressed later in the thesis.

3.3 Users' understanding of YouTube algorithms' inner workings

Before discussing users' opinions on data inputs used by YouTube algorithms, they were asked to share their experience of learning about the existence and functioning of YouTube algorithmic recommendation systems.

3.3.1 Users' ways of learning about YouTube algorithms

For the vast majority (75% of respondents), the process of learning about YouTube algorithms was very intuitive and gradual. It looked mostly like an observation of certain trends and patterns on the platform:

"I don't think it was an apparent realization that I made on my own, because it was very natural and gradual. I started using the Internet very young, as all of people from our generation did. In the 90s, there was yet no YouTube, but we already had a close connection with the Internet and could see that our behaviour online can predispose the recommendations we will get later on" (P07, 28, female, MSc student, Brazilian).

These respondents' hypotheses were later supported by external sources such as articles, videos, classes. This allowed them to incorporate their personal experience and observations into a bigger picture of algorithmic functioning not only on YouTube but on all other major online platforms:

"At first, that was definitely just my observation and I was not even sure if that was actually the case that all of the videos that I get recommended are based on my previous behaviour and interaction on the platform. But then it just that everyone started talking about it in the media, there were some scandals related to social media algorithms and of course I noticed some of headers about it in media and read these articles" (P13, 23, female, project manager, Georgian).

For others, an understanding came with practice as a part of their professional activity:

"I noticed it quite practically because…I am doing digital marketing. And, of course, I discovered that I was able to target a lot of people based on their behaviour on YouTube, for example, or on any other Google platform" (P05, 31, male, digital marketing specialist, Belgian).

Only for 25% of respondents, knowledge of YouTube algorithms came from an external source straightaway – school classes, university classes or documentaries:

"I didn't really read about algorithms anywhere before. It's just that they came up in general discussions, either at university, or again, simply online. The Social Dilemma documentary, for example, on Netflix, also talked about YouTube algorithms" (P18, 25, male, MSc student, Scottish/German).

It is worth mentioning that the Social Dilemma documentary that was released on the famous VOD platform – Netflix – in 2020 was named as a source of education on algorithmic curation for several other respondents. This can indicate that such initiatives contribute to educating a general population on otherwise complicated and highly technical topics.

3.3.2 Users' beliefs on YouTube algorithms' influence factors

In this section, we will discuss user beliefs on influence factors (data inputs) that affect YouTube recommendation algorithms. Here, participants were first asked to critically analyse their personal experience with the platform and name influence factors that first came to their minds. They were then asked more suggestive questions that allowed them to expand their opinions further. This way, the researcher was able to detect which data inputs appear to be more natural for lay users.

As a result, twelve factors defining what will be shown in users' recommended content feed emerged. These factors were then grouped by the researcher into four categories: (1) previous actions or interactions with YouTube (2) video metadata

(3) personal and socio-demographic data (4) external data or data from outside YouTube.

3.3.2.1 Previous actions or interactions on YouTube

This first category of user beliefs on YouTube algorithms involved users' online behaviour patterns on the platform. More than half of all influence factors mentioned were attributed to this category. Previous actions performed online were perceived to be an explicit signal of user's interests and preferences.

3.3.2.1.1 Watch history and clicks

Watch history and/or **clicks** seemed to be the most obvious influence factors and were mentioned by 55% of respondents. In other words, they believed that the content they previously clicked on and/or watched would affect their future recommendations:

"I would say, that might be your clicks. That will allow them to build a profile on you with the details about your interests, preferences, etc. And then they can already start recommending you videos from the same categories or maybe the videos users like you might have watched at a certain point" (P16, 30, female, MSc student, Russian).

Here, it is worth mentioning that more than half of respondents believed that there is a sort of user clustering on YouTube (related to collaborative filtering), meaning that the platform's algorithms are seeking to find connections between users' behaviour on the platform to then use these insights for content recommendation:

"[there are] specific buckets of users that vary depending on how these users behave on the platform, what they engage with and what not, what comment on etc. For example, an art video user will end up in a bucket of users with the same tastes. But you can also be in several buckets and it doesn't mean that you will be allocated to only one bucket and that is it" (P05, 31, male, digital marketing specialist, Belgian).

There were also some (25%) who perceived it as a rather simplistic and outdated algorithmic tool, therefore, denying an existence of such practice on YouTube. This shows a lack of knowledge about hybrid filtering employed by YouTube among some participants. The main explanation that was mentioned several times was users' belief that YouTube algorithms are very personalized and are working on a case-by-case basis:

"Well I guess that could certainly happen earlier, when the technique wasn't still that well-developed. During these early stages, I think they would just scan the users' behaviour, find the ones that are more or less the same and then recommend

them the same types of videos. But I think now when the technology is much further developed, it is not the case anymore, they can really check each user one by one and then suggest a very personalized content" (P11, 25, male, MSc student, German).

3.3.2.1.2 Time spent watching a video

The next major influence factor within the interaction-based category was the **time spent watching a particular video** (named by 55% of respondents). The reason why users believed the number of seconds watched matters for YouTube algorithms was two-fold. Particularly, one group of participants was proclaiming that time spent watching a video gets considered by the algorithms simply because it reflects user's unique online behaviour:

"So, if someone keeps watching a five-minute video, it would be not so good to suggest to them a video that is one hour long, because that's not their usual behaviour, it doesn't fit what they previously did" (P02, 26, female, MSc student, German).

Another argument was that time users spend watching a particular video correlates to time they will spend on the platform. According to participants, the ultimate aim for YouTube is to keep users engaged for as long as possible, and therefore it makes sense that algorithms would recommend content that will keep users on the platform for an extended period of time:

"I would guess they definitely analyse the amount of time I spend watching a particular video. As if I watch a video for only 5 seconds, that might mean I am not that interested in this kind of content. Therefore, when I watch a video for even 10 second, they might tell that this video got me involved for longer" (P07, 28, female, MSc student, Brazilian).

Respondents who did not mention this influence factor themselves were asked whether they believed that time users spend watching a video influences YouTube algorithm in one way or another. Noticeably, all of them agreed.

The opposite behaviour, namely, **skipping the video** after watching it for a brief period of time, was also seen to be a valuable algorithmic input. The algorithm was believed to be down-ranking the videos that users scrolled through as such action was likely to be a result of a mistake or an indication of pure disinterest:

"…maybe I just turned on the wrong video, I didn't watch it at all and just turned it off. Then that means that such video will not really correspond to my needs and expectations. Therefore, I don't think they would still recommend it later on" (P13, 23, female, project manager, Georgian).

3.3.2.1.3 Likes, dislikes, comments

Likes, **dislikes** and **comments** were also deemed a valuable contributor to YouTube algorithms' functioning by many respondents (45%). These influence factors were considered a strong signal from users, and were therefore seen to have a major influence on their future recommendations.

While the first two factors – likes and dislikes – were thought of by the majority of the participants straightaway, comments' relevance was questioned by some of them:

"And if you like or comment on certain videos, it will also include that in the algorithm, I would say. I mentioned comment section but I'm not too sure about whether the fact that you have commented on something is as valuable as like or dislike" (P03, 21, male, MSc student, Belgian).

In some instances, respondents did not mention this influence factor without the researcher's assistance. One of the explanations for ignorance towards comments as input for YouTube algorithms is users' intuitive understanding of algorithmic functioning on YouTube which could also be referred to the concept of user beliefs that are purely based on users' previous interaction with the system (Gelman & Legare, 2011). Therefore, as their personal engagement with the platform often excluded leaving comments on the videos, some respondents were struggling to express their opinion on this question:

"Yeah, it's kind of complicated question for me because I don't comment that often. But, I guess, again, there is kind of filter that the checks everything and for sure notes down whether I have commented on something" (P16, 30, female, MSc student, Russian).

To allow participants elaborate more, they were asked whether they think YouTube algorithms were analysing the content of the comments. While some interviewees agreed that sentiment analysis performed by AI could be used to extract insights from comments, the vast majority suggested that it is not the emotion users experience that matters but users' engagement:

"I think they have an ability to analyse the content of the comments but they don't really use it for their recommendations. Any reaction is good for them, it's like, if a video gets a lot of dislikes, it will be still recommended to a lot of people because it creates a lot of interaction" (P20, 32, male, product owner, Belgian).

Interestingly, another form of interaction such as **sharing** a video was not mentioned by any of the participants until the researcher asked a related question during the suggestive part of the interview. An overwhelming majority of respondents (80%) agreed that such action will most likely result in similar videos to be listed in the recommendations section on YouTube:

"I think it would be quite logical because it gives the insights on users' preferences and true interests. Without enjoying the video, I think, you will not share it with your friends" (P13, 23, female, project manager, Georgian).

However, some interviewees questioned whether it is the same for all the social media platforms and drew a distinction between publicly available content and something shared in private messages:

"Yeah, there is cross site tracking, for sure. I mean, I'm not sure about whether it's happening on WhatsApp because it is encrypted. But if I will, for example, share it on Facebook publicly, then for sure they would see it and consider for future recommendations" (P11, 25, male, MSc student, German).

Here, certain privacy-related concerns were touched upon by some users for the first time:

"I don't know if they know, if they can tell every time someone's shared, even in private messages a specific video. I would say 'no' because it feels like it would be a violation of privacy to be able to know what I write in my personal messages, even if what I write is a link to their video" (P17, 30, male, music producer, Belgian).

3.3.2.1.4 Subscriptions

In addition to the previous interaction-based influence factors, the **channels people are subscribed to** were also deemed as indicators for similar videos or videos watched by users with the similar preferences to appear in the recommendations' feed:

"I guess the first would be the accounts I follow because at least half of the recommendations I get are from these channels" (P10, 24, female, MSc student, Ukrainian).

Even users whose usual behaviour on the platform did not necessarily include subscribing to any channels agreed during suggestive interview part that subscriptions have a definite value for algorithms:

"I don't have any subscriptions on YouTube actually. That is why I have no idea how that works. But I am quite sure that makes sense because when you subscribe, you click on this little bell then you are already sending the signal to the platform that you want more of that content in your feed" (P11, 25, male, MSc student, German).

It was also suggested that subscription allow YouTube to promote users' engagement with the platform by sending notifications about new content coming from the channels one is subscribed to:

"…when you subscribe, they usually suggest to enable the notifications for the other videos from this channel. It should really be related because also I think that the

fact you are subscribed to a particular channel might show your spectrum of interests and preferences" (P19, 27, female, MSc student, Italian).

3.3.2.1.5 Search queries or keywords

Finally, another major influencer within this category were **search queries**:

"One thing certainly is the use of trending tag words that you put into the search bar, so, for example, when I type in 'Trump versus Hillary Clinton', it will compare my search query with videos themselves as well as to videos that people with the similar search history have been looking up and therefore suggest me videos that are therefore relevant" (P03, 21, male, MSc student, Belgian).

Mostly, those suggesting that keywords act as one of the primary influence factors for algorithmic content curation on YouTube were referring to the platform as a search engine rather than a social media platform:

"I think that whenever it comes to search engines to which we might also relate YouTube, keywords or generally something that you type in the description or meta titles is important" (P13, 23, female, project manager, Georgian).

After being asked whether searches without any supportive interaction (clicks) will affect YouTube algorithms in the same way, the majority (65%) of respondents agreed that search results on their own are being stored to use for recommendation purposes later. It is also worth mentioning that an association with YouTube's parent company, Google, was brought up here:

"I think that might be the same as with Google, when you are searching for something, it gets logged in their system to then retarget you for advertisement purposes etc. I think that is how it works with YouTube too" (P06, 24, female, MSc student, Italian).

At the same time, some interviewees addressed the question differently suggesting that even though algorithms will consider it at some point, such action will be weighted less:

"I think the fact that you arrived on specific pages already means something to them. But I think it will be weighted less than if you had clicked on it. So, the degree of importance will be less but it will still be considered by the algorithms" (P05, 31, male, digital marketing specialist, Belgian).

3.3.2.2 Video metadata

YouTube's interface frequently allows users to navigate an array of video content and filter it out by looking at others' activity, which can be visualized through videos' metadata: number of views, video's description, title, thumbnail.

In this section, we will learn how users associated these features with YouTube's algorithmic functioning.

3.3.2.2.1 Popularity factor or view count

It was agreed that the overall **popularity** of the video can influence YouTube algorithms one way or another:

"I think it is important for algorithms to show users something that the rest of community on the platform has engaged with" (P05, 31, male, digital marketing specialist, Belgian).

However, it was mentioned that regardless of how popular a certain video is among others, it will not be shown in the recommendations section if it does not have any relevance for a user:

"Of course, what is now trending and regarded as popular will be recommended much more. But I mean that the most popular video from the category of videos I already watch would be recommended, not just any random video with lots of views on it" (P11, 25, male, MSc student, German).

Such claims once again suggested that users generally believe that YouTube algorithms are tailored for the personal preferences and interests of each user, and aimed at improving their user experience.

3.3.2.2.2 Video's description, title, thumbnail

Furthermore, keywords used in the **description**, **title** and the choice of a **thumbnail** were also seen to be a relevant contributor to the functioning of algorithmic suggestions on YouTube:

"Good description with a lot of keywords words can be taken into account by YouTube algorithms, or at least I think so" (P15, 22, female, MSc student, Belgian).

Even though the majority of respondents (70%) when asked a related question agreed that description, title, and thumbnail are a valuable influence factor, their reasoning behind it varied.

Namely, once again a parallel with the search engine was made by some to explain how keywords can influence algorithms:

"It is like a search engine. So, if I am typing in the word "nature" in the search box the best video, with the best narrative and quality is less likely to get enough views to appear in the search results than a poorer-quality video with this keyword in a description and the title" (P01, 23, female, business manager, Georgian).

Meanwhile, others insisted that eye-catching and organic keywords will attract users' attention which will then influence algorithms respectively:

"The title cannot influence on algorithm, but it can influence the users' behaviour – rather attract users' attention or not. And that is what then will influence the algorithms" (P16, 30, female, MSc student, Russian).

It is also worth mentioning that, because of their lack of knowledge on how keywords in description and title can affect YouTube algorithms, some users were trying to compensate with their insights about other platforms:

"I do think that title and the image thumbnail you choose will have an influence. I'm not sure though if that would do the trick at all actually. Because I know that in case with Facebook it is not a good idea to change something about the post once it is published, will affect badly the algorithms" (P10, 24, female, MSc student, Ukrainian).

3.3.2.3 Users' personal and socio-demographic data

The third category of influence factors contributing to YouTube algorithmic functioning was related to explicit and implicit user data.

3.3.2.3.1 Gender and age

Demographic data such as **gender** and **age** were among one of these factors:

"Well, first of all, probably the demographics, and the information that is provided in my YouTube profile. I'm not sure maybe even gender, my age, some information that I provide about myself when creating an account" (P14, 24, female, MSc student, Lithuanian).

3.3.2.3.2 Location

The algorithms on YouTube were also claimed to be considering information on users' **location**:

"…if I am changing the VPN sometimes on my laptop, I have different kind of videos in my suggested videos. I'm from Georgia, therefore, I will mostly get the videos about Georgian culture or by uploaded by Georgian YouTubers but if I'm on the platform with my American VPN, then I will have a completely different type of content suggested in the Trending Now" (P01, 23, female, business manager, Georgian).

Notably, some of the respondents noticed that location would be only considered if the location settings are enabled on a device:

"Depends on whether you enable the location access for the app in your setting. But because I haven't done that, I got more kind of global stuff" (P09, 26, female, MSc student, Swiss).

Advertisement on the platform was also mentioned here:

"I think for the advertisement, yes, the location gets considered. Because when I was in Brazil, I used to watch some local news, as I mentioned already. So, they kept recommending me the same content, but the advertisement in the video was in Dutch, or in French" (P12, 25, female, MSc student, Brazilian).

3.3.2.3.3 Time of the day

Finally, the **time of the day** was also considered to be one of the algorithmic contributors:

"I guess I would rank it [influence factors] like 80% are really coming from my user behaviour like clicks or even my searches, but I'm guessing that around 20% could be influenced by my surroundings and my location, as well as the time of the day. I think I might have different recommendations at the evening than I would get in the early mornings" (P11, 25, male, MSc student, German).

Even though time of the day is dependent on a context of utilization, as it is closely linked to the location of user, it was considered to be one of the demographic variables of users' data.

3.3.2.4 External sources or users' online activity outside of YouTube

Even though, generally speaking, users were not concerned about YouTube getting any of the insights from their online activity outside of the platform, there were some (15%) who still claimed that there is a connection between algorithmic functioning on YouTube and **Google searches** or **online shopping**:

"…sometimes I see that…I can be doing online shopping on some websites and then end up on YouTube, there, I might see a video of an influencer presenting a certain brand" (P06, 24, female, MSc student, Italian).

However, it is worth clarifying that users distinguished Google products and those of other companies. Therefore, when asked about whether sharing videos to other platforms would mean that they would get similar videos recommended later, those who were denying such opportunity mainly were referring to the following argument:

"That would mean that Facebook and Google would have to work together and share information among each other, I don't believe that would be possible in today's reality" (P01, 23, female, business manager, Georgian).

But the possibility of sharing users' data such as their interests and preferences between two Alphabet Inc. subsidiaries – Google and YouTube – was also clarified:

"If you have a YouTube account, you put some data in there or maybe simply connect it with your Google, with Gmail" (P02, 26, female, MSc student, German).

3.4 Conclusion on users' understanding of YouTube algorithms' inner workings

Before moving to the second part of the result discussion, it is important to draw some conclusions on users' understanding of YouTube's algorithmic functioning.

The Fig. 5 below shows the type of data respondents believed to be used by YouTube to feed their recommendation algorithms:

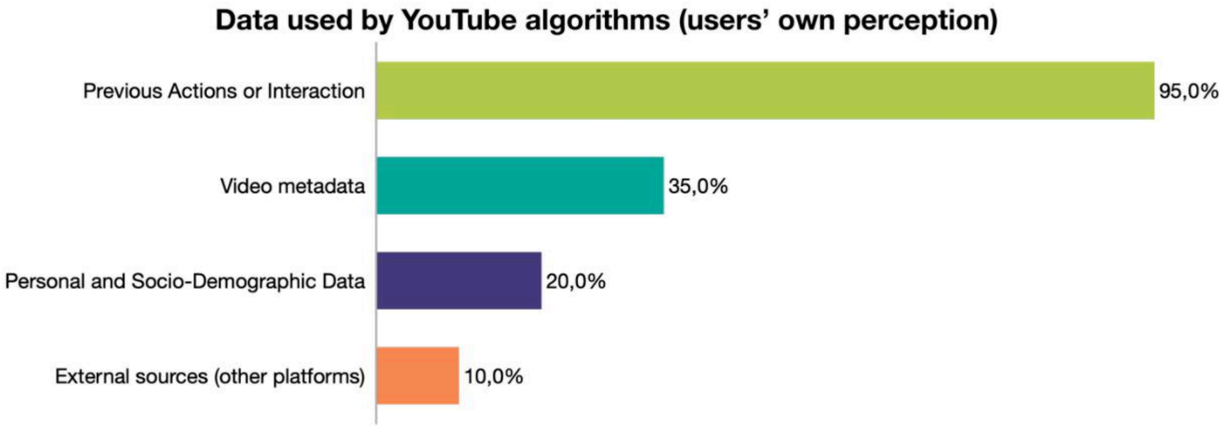

Figure 5: Data used by YouTube algorithms (users' own perception)

By looking at the figure above, we can notice that respondents have mostly prioritized their previous actions or interaction with YouTube as the main influence factor. This might be explained by a lack of technical understanding among users and, thus, a tendency to base their reasonings on their personal experience with the platform (user belief).

Another reason might be coming from users' belief that algorithms on YouTube are very sophisticated and are precisely suited for each individual.

Data such as watch history, clicks, search queries, likes, dislikes, comments, shares, subscriptions, time spent watching a video were often mentioned as key influencers on YouTube recommendation systems.

The second most-common influence factor category was video metadata. This category implies that other users' actions can also affect recommendations. It explains why even though none of the participants initially mentioned user clustering as one of the influence factors, the majority of them agreed that it still might play a significant role once they had been presented with a suggestive question on the topic.

Finally, personal socio-demographic data and activity outside YouTube were hardly seen as important as video metadata. Only 15% of all respondents reported that YouTube draws data from affiliated external sources (Google searches and online shopping). While these users recognized the existence of free data movement between platforms related to YouTube, they have never mentioned the possibility of YouTube gaining access to their data from platforms and websites that are not directly linked to it – Facebook, Instagram, Twitter. This shows a certain level of trust towards the way YouTube utilises its users' data.

Overall, the vast majority of participants showed a high yet superficial level of awareness about YouTube's algorithmic functioning. While they were aware of its existence, their understanding of the algorithm's inner workings was noticed to be limited and vague, mostly based on their own observations. Almost all of the data inputs (except for subscriptions) appeared to be "natural" for respondents as they mentioned them without researcher's assistance.

It was also observed that participants, in some instances, held more in-depth comprehension of algorithmic functioning from other SNSs such as Facebook and/or Instagram, and were trying to link it to YouTube. This, however, could be explained by recent recurrent scandals related Facebook and its subsidiaries (i.e. Cambridge Analytica scandal in 2018) which received an extensive media coverage.

3.5 Users' attitudes towards algorithmic curation on YouTube

When discussing possible advantages and downsides of algorithmic curation on YouTube, the predominant majority of participants (75%) expressed their strong support for YouTube recommendations. The added value they provide to users' experience on the platform was pointed out to be one of the algorithms' eminent advantages:

"So, I personally think about the recommendation systems as something beneficial because most of the times they improve my experience with the platform, that makes me wanting to find content that I like. I know that algorithms can help me in that" (P03, 21, male, MSc student, Belgian).

Even though most participants showed understanding that algorithms are employed for profit maximization and monetization, they admitted that they were enjoying using the platform and will continue doing so as long as they keep getting engaging and/or relevant content:

"I think that YouTube as any platform nowadays needs this cross-side network effect to be running. So, they need to keep me engaged. But at the same time, they want YouTubers to make content and advertisers to pay for ads on the platform. So, I kind of understand this need for algorithms, therefore, I don't have anything against them, for me, it is more positive" (P10, 24, female, MSc student, Ukrainian).

There were also some users who expressed a high level of trust in YouTube's altruistic motivations:

"I believe that people working in YouTube are really there to propose new content, to make people discovering stuff because honestly, I personally learned a lot of stuff because of YouTube. So, people working there they don't have this evil intention, we can say that they are actually working for the benefit of users" (P20, 32, male, product owner, Belgian).

Several participants raised the question of filter bubbles and echo chambers. While they believed their high level of digital literacy protected them from these biases, they feared that less educated users were under threat of being steered towards radical views and content. More than a half of respondents were convinced there are certain societal concerns related to these mechanisms we cannot ignore:

"My concerns are mostly related to filter bubbles, echo chambers. I mean, when I look at me, I don't have the threat that I will get radicalized online. Because I'm socially well-integrated, have good education, some understanding of algorithms, but I'm sure that sometimes the YouTube algorithms can lead towards radicalization of users that might not have a very well-shaped understanding of the world" (P09, 26, female, MSc student, Swiss).

This illustrated the need for the promotion of media and digital literacy among a broader audience, especially one that is more vulnerable to algorithmic influence.

Improving media literacy among those who have no or minimal understanding of algorithmic curation could resolve a power asymmetry issue and develop an audience that would be more independent in shaping its opinion.

Interestingly, while privacy concerns related to social media platforms have gained strong momentum during the last few years, respondents have not explicitly shown any concern about the safety of their data on YouTube. This attitude could be explained by the benefits offered by YouTube algorithms, low sensitivity of data stored on YouTube, as well users' perception of the platform as a search engine rather than a social medium:

"Well I think I feel less concerned with my personal data security and I don't have as many privacy related concerns when it comes to YouTube but I am not sure why. I might guess that it is because I don't communicate with anyone really on YouTube, I don't upload anything there either" (P02, 26, female, MSc student, German).

Another explanation for users' seemingly carefree attitude towards data protection on YouTube could be linked to a 'black boxed' nature of YouTube algorithms which is likely to deprive users of an ability to evaluate the threats related to algorithmic curation (Pasquale, 2015).

Some, however, explained the lack of privacy concerns in relation to YouTube by claiming that the platform employed strong data security practices:

"I see that situation with the personal data security as putting money in the bank. For instance, you can keep your money under the pillow or you can deposit them in the bank. I tend to think that in the bank they will be safer after all, the chances that someone will steal your money are almost zero. I know for sure that people working in the IT department for Google are the best of a kind, therefore, their data security is on the top level" (P05, 31, male, digital marketing specialist, Belgian).

By looking at the results of data analysis, we can conclude that the platforms' benefits could outweigh the downsides users mentioned. These findings echoed the results of Fast's and Jago's study (2020) that showed that the absence of motivation to protect their privacy online, among other factors, might be coming from benefits provided by the algorithms and no expectancy of serious harm (Fast & Jago, 2020).

According to Taylor (2004), privacy concerns are often context-specific, and this hypothesis was supported by the results of our observations (Taylor, 2004). Indeed, while respondents have shown their understanding of potential data-related concerns often associated with other social media platforms, they did not feel the same way about YouTube.

3.5.1 Empowerment and feeling in control

When discussing users' empowerment and control over recommendations they get on YouTube, the predominant majority (65%) of participants expressed a high level of agency on the platform. These findings showed opposite results in comparison to other more communication-oriented social media platforms as Rader and Gray (2015) found that only a small fraction of their sample felt empowered and in control of their Facebook News Feed (Rader & Gray, 2015). This could be partly explained by YouTube's primarily leisure-oriented purpose, as well as users' seemingly high level of awareness about algorithmic functioning on YouTube.

The reasonings behind our respondents' opinion varied. While some attributed it to their specific use of YouTube:

"I feel in control because, as I said, I rarely simply click on the videos that get recommended to me on the home page. I simply search for something. Therefore, I can just change my search and then the recommendation algorithms will adapt accordingly" (P01, 23, female, business manager, Georgian).

Others associated it with YouTube's algorithms' adaptive nature:

"Yes, because, for example, now the tutorials for music production I am watching are not going to be there if all of the sudden I will start watching videos about butterflies. In a week max the recommendation algorithms will adapt to my behaviour on the platform" (P20, 32, male, product owner, Belgian).

Meanwhile, some participants recognized that their control over YouTube recommendations was relative and ambiguous:

"I think we have certain control over them, but not too much because they YouTube algorithms will adapt to whichever content I am watching. Even though I will try to 'fight' against it, they will follow me and start recommending me this new type of content after some time...Feels like we can't escape the algorithms, they're always one step ahead and that's kinda scary" (P12, 25, female, MSc student, Brazilian).

Although, there were still a few participants (10%) who admitted that the constant surveillance and feeling of intrusiveness deprives them of their agency while they are on YouTube:

"For sure I don't really feel in control because I know there is something that watches over me all the time I am on the platform. Of course, this is not a real human-being but a machine, but still, I don't feel very comfortable and safe…we can try and change the recommendations we get, but this constant feeling of being a 'puppet' will never disappear" (P16, 30, female, MSc student, Russian).

At the same time, some participants used the same argument of algorithmic surveillance in support of the platform saying that they feel less worried knowing that machines are analysing their behaviour online and not actual human-beings.

3.5.2 Curated vs. Non-curated YouTube feed

When asked about whether they would choose curated or non-curated YouTube feed, half of respondents expressed their preference towards the first option. Considering all pros and cons discussed, participants confirmed that they are fully happy with algorithmic curation on YouTube because of the added value it brings to their user experience:

"I would say that the user experience will be better with the recommendation systems because that means that they don't have to try to search for something and also there is a chance that these recommendations will suggest something new and exciting you would not have found otherwise. So, I personally would choose the

platform with the recommendations but yet being aware of possible constraints" (P05, 31, male, digital marketing specialist, Belgian).

While half of the participants confirmed they would not switch to an alternative platform with no content curation even if given a choice, 40% of interviewees confessed they would appreciate to at least have a choice between these two options and be granted some freedom in managing their YouTube recommendations:

"Yes, I think the last one [having a choice between the two]. Because then users can have a kind of feeling that they can decide something. If they [users] do not want to be analysed, I mean their digital behaviour, their footprint, they can choose an option of a platform without any algorithms on it" (P16, 30, female, MSc student, Russian).

Meanwhile, a small fraction of participants (10%) was open to an alternative version of the platform with no content curation. Mainly, this decision was explained by users' desire to avoid distraction and gain more control over time spent online:

"I think I would be happy with no suggestions. Because if I want to watch a certain thing I just need to look for it, I don't need the recommendations. I think I would enjoy more freedom on the platform. I would have less distractions without recommendation and spend less time on YouTube, that is for sure" (P12, 25, female, MSc student, Brazilian).

We can summarize that even though participants demonstrated an understanding of revenue-oriented rationale behind YouTube recommendation systems, as well as concerns related to their functioning, they mainly expressed a high level of satisfaction with the quality and diversity of recommended content on YouTube.

3.6 Users' attitudes towards algorithmic transparency on YouTube

A great majority of respondents (70%) were convinced that there is an apparent need for more algorithmic transparency on YouTube. At the same time, they expressed their awareness of possible motivations for YouTube to keep the functioning of their algorithms unavailable for the general public. Their opinions varied on whether YouTube should be entitled to keep the nuances of its algorithmic functioning unexplained.

Here, some participants, while admitting an apparent need for transparency, have also supported YouTube in its right to preserve its intellectual property and competitive advantage by not disclosing details of its algorithms' inner workings:

"I think they have the right to say it's a business matter and they are in their own right to keep their competitive advantage they were developing for ages for themselves. So, if supporting a profit-oriented approach, it is possible to defend the platform. But as user we have also the right to know what's behind the service we use so much" (P05, 31, male, digital marketing specialist, Belgian).

The same ideas appeared in academic publications by Diakopoulos and Koliska (2011), as well as Diakopoulos (2016), that highlighted a lack of business incentive for big online platforms to reveal their algorithmic functioning logic (Diakopoulos & Koliska, 2017; Diakopoulos, 2016).

However, commercial interests were not the only explanation for the lack of algorithmic transparency on YouTube. Several respondents agreed that this decision might be coming from the platform's desire to preserve content quality and prevent any bias:

"For me, the platform is right not to reveal how their algorithms work for a different reason. If content creator will know how it works, they will not be eager to produce a good quality content they will learn the rules of the game and boost their videos even if they are not interesting and of bad quality in general" (P20, 32, male, product owner, Belgian).

Additionally, it was pointed out that transparency is not absolute and while regular users might not be interested or not be able to comprehend all the technical insights of YouTube algorithms' inner workings, certain stakeholders should be eligible to get a hold of this information:

"…for regular users I don't see a need or a demand for knowing such technical details. Even for us [communication studies students] now it is mind-blowing sometimes and it something that is really hard to comprehend" (P09, 26, female, MSc student, Swiss).

"I think it should be case-by-case. For example, it is important for YouTubers trying to make money on the platform. Or it might be even interesting to have a full information on how the algorithms work for governmental bodies which are trying to monitor and regulate platforms like YouTube" (P10, 24, female, MSc student, Ukrainian).

These opinions can support findings of Diakopoulos and Koliska (2017) stating that a complete algorithmic transparency might not be a necessity when it comes to lay users as it will be challenging for them to comprehend technical insights of these mechanisms (Diakopoulos & Koliska, 2017). A similar conclusion was also presented by Beer (2017) who claimed that due to the very complicated technical nature of algorithmic functioning, it is overwhelmingly hard for lay users to fathom all aspects of the question (Beer, 2017).

At the same time, several respondents expressed their strong conviction that users are entitled to know how their data helps YouTube algorithms to generate recommendations:

"…it must be transparent to the users, it needs to be clear about what data they are collecting, there must be a way to delete your personal data from the platform, or at least to look into what data you have submitted" (P03, 21, male, MSc student, Belgian).

Some respondents pointed out that a lack of transparency online, in general, is a problem of many hands and cannot be solved by any of the stakeholders alone. That is why a corporate initiative is needed:

"I think everyone has to put effort, we can't have just a single stakeholder that would be responsible for a whole situation. It is a many hands problem, therefore, we need some corporate actions here" (P15, 22, female, MSc student, Belgian).

Meanwhile, there were also several participants (30%) who did not express any concern regarding a lack of transparency on YouTube:

"I don't really care about it that much to be honest. I've never had any bad experience with recommendation systems myself because usually what get recommended is exactly what I have been looking for" (P08, 24, female, MSc student, Polish).

3.7 Responsibility for users' algorithmic education

A great majority (85%) of participants expressed their profound interest in learning more about YouTube algorithms' functioning. However, they had divergent opinions about who should be responsible for users' education on the topic (see Fig. 6).

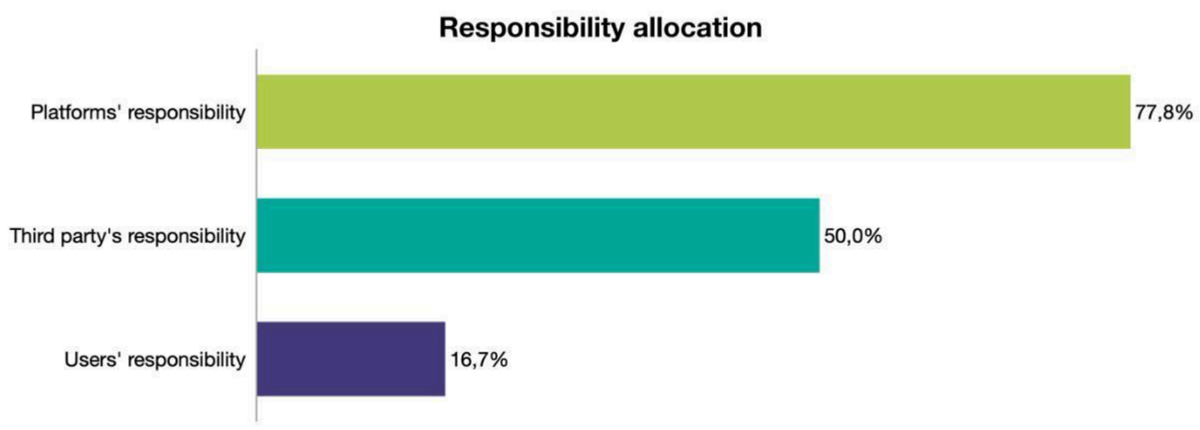

Figure 6: Responsibility for users' education on YouTube's algorithmic functioning

As the figure hereinabove shows, 77.8% of interviewees considered that it should be the platform's responsibility:

"Of course, I would like to know more. Because as I said, sometimes I don't really understand why I'm recommended one specific thing rather than another. I am guessing it must be advertisement but it could be something completely different. I might do something on the platform just because I am not aware of how it actually functions. It is like with accepting cookies all over the Internet, sometimes they might be asking for your fingerprint and we still accept it just because we want to get rid of this pop-up window. So, yes, I think that the platform should be more transparent with that and it should be mostly their responsibility to educate users because they are the ones who created the algorithms in the first place" (P06, 24, female, MSc student, Italian).

Some respondents (15%) claimed that users themselves should also be more proactive in educating themselves on this topic:

"I feel that, of course, the platforms are responsible and that they need to be more responsible…But then users should also be more aware of privacy rules and should think about educating themselves on such topics as much as they can" (P15, 22, female, MSc student, Belgian).

Furthermore, it was suggested that third-party organizations, independent agencies, governmental bodies, education institutions should uptake the responsibility of opening up online platforms for public inspection and promote media and digital literacy among users:

"I think an initiative with the schools providing some classes on this topic would work well because it has worked before with fake news. I mean there was this incredible initiative in Sweden about the fake news, Facts vs. Fake I think it was called. So, educating young people would be great but by doing that we don't totally solve the problem because we have older users who don't even realize what's happening" (P07, 28, female, MSc student, Brazilian).

"I always think it's biased when it's coming from the platform itself, that they select what they show and what not. I like the political attempt to force the big platforms

to open, like to be more transparent or to be entirely transparent" (P09, 26, female, MSc student, Swiss).

"I think it has to be some sort of NGO, or even the government. So that we would be reassured that there will be enough money for research, they would be transparent with who is funding them and they would be qualified enough to teach the audience about such a topic. Or it could be a documentary like Social Dilemma. So, third party" (P10, 24, female, MSc student, Ukrainian).

These remarks showed an ultimate interest among users in learning more about algorithms on YouTube. While there are some who admitted that the technical complexity of the question might cause more confusion among lay users, most participants expressed a need for more transparency and users' education on this topic.

4. SUMMARY

With the present study, we have investigated young users' beliefs about YouTube recommendation algorithms and looked into their meaning-making of algorithmic functioning on the platform. We also examined young users' empowerment and their attitudes towards YouTube algorithmic content curation and transparency.

Probing users' awareness of a topic such as YouTube recommendation algorithms was tricky due to its technical complexity and interviewing process dictated by COVID-19 restrictions. Therefore, open interview questions were followed up with suggestive questions, which noticeably expanded respondents' outlook and allowed them to elaborate on certain topics.

A qualitative analysis of users' opinions about YouTube algorithms' data inputs and their attitude towards algorithmic content curation on the platform allowed us to highlight several valuable observations.

The first research question was targeted at examining how users learned about the existence of YouTube algorithms in the first place. Generally, the learning process for participants was found to be intuitive, gradual, and deeply rooted in their personal experience with YouTube. Schools, universities, or other external sources were mentioned by only a few of our respondents (25%), implying a lack of initiative from educational and governmental institutions to enlighten citizens on algorithmic curation and how it shapes their common sense. At the same time, informal initiatives, such as the Social Dilemma documentary, were seen to contribute significantly to the population's education on otherwise technically advanced questions.

When probing into participants' beliefs on data inputs employed by YouTube algorithms to provide recommendations, our users recognized a diverse range of influence factors. The researcher then grouped these factors into four categories:
(1) previous actions or interaction with YouTube (2) video metadata (3)
personal and socio-demographic data (4) data gathered from outside YouTube.

Data collected from users' previous interactions with content on the platform was believed to be the most influential factor (95%) for YouTube recommendation algorithms. Data inputs such as users' subscriptions, watch history, time spent watching a video, skipping a video, and search queries were mentioned in this users' beliefs category.

Actions deemed to decide whether certain content will appear in one's YouTube recommendations also included liking, disliking and commenting on a video. While likes and dislikes appeared to be obvious data inputs, comments' relevance was questioned, which illustrates how user beliefs are being shaped. Indeed, since commenting on videos was not part of users' normal behaviour on the platform (meaning that it was excluded from user beliefs), they struggled to explain how that could affect algorithms.

These findings suggest that YouTube recommendation algorithms are considered a sophisticated socio-technical system capable of analysing each user's personal experience with the platform to provide relevant content propositions.

The second-largest category of influence factors (35%) consisted of video's view count, as well as keywords in title, description, and a choice of a thumbnail. These results revealed that YouTube algorithms are believed to be evaluating both the video's potential attractiveness for each individual and the content's popularity among other users on the platform.

However, it is worth mentioning that participants noticed that regardless of how popular the video is, it will not be recommended unless it suits the user's preferences. This demonstrates that the majority of users trust YouTube to prioritise content recommendation's relevancy over popularity.

Thirdly, only a tiny fraction of respondents (20%) mentioned personal and socio-demographic data such as users' gender, age, location and time of the day.

Finally, a minority of respondents (10%) claimed that YouTube recommendation algorithms were extracting data inputs from outside the platform. However, they believed this data collection process only happens between the platforms affiliated with YouTube or its parent company, Google. None of the participants expressed the possibility of YouTube gathering and using data from social media platforms and websites that are not directly linked to them. This, in turn, can be seen as an indicator of users' trust in the way YouTube utilises their data.

Overall, it was found that while respondents held a high level of awareness of YouTube recommendation algorithms and their data inputs, their understanding of the algorithm's inner workings was limited and vague, primarily based on their observations and intuition.

Although users have mainly (65% of respondents) claimed to hold a high level of agency over the content recommended on YouTube, their reasoning behind it varied. Some users were convinced that the highly adaptive nature of YouTube algorithms and advanced platform settings allowed them to steer recommendations towards their preferences. Meanwhile, others recognized that their control over YouTube recommendations is relative and ambiguous as they are contributing to algorithmic functioning on the platform one way or another. There were also some (10%) who admitted their lack of agency on the platform due to being constantly observed and analysed.

While expressing awareness of the business rationale behind YouTube recommendation systems, users have shown a positive attitude towards algorithmically curated YouTube recommendations, satisfaction with the quality of recommended content and a certain level of trust with the platform's data-sharing practices. Half the respondents claimed that they would not switch to an alternative non-curated platform even if given a chance. Meanwhile, many respondents (40%) have admitted that they would appreciate having a choice and being given more freedom over their recommendations on the platform. Only a minority of respondents (10%) acknowledged they would like more agency and control over their video content consumption and, if given a choice, would switch to a non-curated YouTube feed.

Half of the participants expressed concerns over filter bubbles and echo chambers. While they believed their high level of digital literacy protected them from these biases, it was deemed essential to promote media and digital literacy among those who are more vulnerable to algorithmic influence.

Interestingly, no privacy concerns were brought up in relation to YouTube algorithms' operation. This attitude, however, could not be linked to users' lack of understanding of the subject. Instead, it could be explained by the low sensitivity of data stored on the platform, benefits provided by algorithmic content curation, and users' understanding of YouTube as a search engine instead of a social medium.

While discussing algorithmic transparency, some participants brought up a conflict of interests between YouTube's financial motivations and the public good. The business rationale was named one of the possible motivations for keeping details on algorithmic functioning undisclosed. However, financial interest was not the only explanation for the lack of algorithmic transparency. Some suggested that the decision to keep YouTube's business leverage a secret might be coming from the platform's desire to preserve content quality and prevent stakeholders, mainly YouTubers, from finding loops in the systems to promote their own content.

Even though most respondents prioritized transparency, quite a few noticed that while regular users might not be interested in learning how these socio-technological systems are functioning, regulators and analysts must be able to access these insights. Respondents' opinions resembled the idea that while transparency is essential, it is often context-specific and should only be made available to certain stakeholders. Thus, another key takeaway is the perception that full transparency might not be necessary or even beneficial for lay users due to the technical complexity of algorithmic functioning on YouTube and other SNSs.

The predominant majority of respondents expressed their profound interest in learning more about the functioning of YouTube algorithms. However, the responsibility for users' enlightenment was allocated differently. Firstly, it was declared that platforms such as YouTube should be responsible for users' education on data inputs and algorithms' functioning. Secondly, a need for a more proactive behaviour coming from users themselves was pointed out. Even though a high level of trust was expressed towards YouTube, its reliability in providing insights into its business model was doubted. Thus, to avoid bias, third-party agencies and research centres were also believed to be an alternative source of information in this regard. At the same time, a few respondents referred to the dilemma as a "problem of many hands". They were convinced that a more corporate initiative is required to spread algorithmic literacy among citizens and obtain more transparency in the functioning of YouTube recommendation algorithms.

4.1 Limitations

A few challenges for conducting research were recognized throughout the process. Firstly, the interviews were conducted while strict COVID safety measures were still in place, meaning that face-to-face communication with respondents was impossible. Therefore, online real-time interviews were employed instead, which could have limited a sense of comfort and trust between the researcher and respondents.

As within this research, participants were aged between 18 and 35 years old, the generalizability of the research to other age groups, primarily those under 18 years old, was limited. The vast majority of respondents were students of the Communication Studies faculty, meaning that they were already aware of YouTube algorithms before the research. Therefore, a certain sample homogeneity limited the generalizability to those who have no prior awareness of the topic.

As observations showed, some participants shaped their user beliefs based on their interaction with the platform: liking and sharing videos, leaving comments, etc. It could be of interest to other researchers to study how users' behaviour online and their interaction with particular platform features can influence their experience, user beliefs, and attitudes.

While this study has probed into YouTube's recommendation systems, further research could be expanded to other video-sharing or streaming platforms.

Moreover, user beliefs could be examined quantitatively with a non-purposive sampling, which could considerably increase the research scope and allow for more generalizability of results. That, however, might include additional efforts from the researcher to develop a questionnaire that would address this technical topic with the employment of appropriate and understandable language.

REFERENCE LIST

The bibliography of this Master's thesis is divided in two parts: academic and non-academic literature.

ACADEMIC LITERATURE

- Abbas, M., Riaz, M. U., Rauf, A., Khan, M. T., & Khalid, S. (2017). Context-aware YouTube recommender system. In *2017 International Conference on Information and Communication Technologies (ICICT)*, 161-164. IEEE. https://doi.org/10.1109/ICICT.2017.8320183
- Aditya, P. H., Budi, I., & Munajat, Q. (2016). A comparative analysis of memory-based and model-based collaborative filtering on the implementation of recommender system for E-commerce in Indonesia: A case study PT X. In *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 303-308. IEEE. https://doi.org/10.1109/ICACSIS.2016.7872755
- Alsop, R., Bertelsen, M., & Holland, J. (2006). *Empowerment in practice: From analysis to implementation*. The World Bank.
- Alvarado, O., & Waern, A. (2018). Towards algorithmic experience: Initial efforts for social media contexts. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1-12. https://doi.org/10.1145/3173574.3173860
- Alvarado, O., Abeele, V. V., Geerts, D., & Verbert, K. (2019). "I really don't know what 'thumbs up' means": Algorithmic Experience in Movie Recommender Algorithms. *In IFIP Conference on Human-Computer Interaction*, 521-541. Springer, Cham. https://doi.org/10.1007/978-3-030-29387-1_30

- Alvarado, O., Heuer, H., Vanden Abeele, V., Breiter, A., & Verbert, K. (2020). Middle-Aged Video Consumers' Beliefs About Algorithmic Recommendations on YouTube. *Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2)*, 1-24. https://doi.org/10.1145/3415192

- Anderson, M., & Jiang, J. (2018). Teens, social media & technology 2018. *Pew Research Center, 31*(2018), 1673-1689.
- Asanov, D. (2011). Algorithms and methods in recommender systems*. Berlin Institute of Technology, Berlin, Germany.*

- Aufderheide, P. (1993). Media Literacy. A Report of the National Leadership Conference on Media Literacy. *Aspen Institute, Communications and Society Program*.

- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science, 26*(10), 1531-1542. https://doi.org/10.1177/0956797615594620
- Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review,* 104(3), 671-732*.* https://doi.org/10.2139/ssrn.2477899
- Basch, C. H., Basch, C. E., Ruggles, K. V., & Hammond, R. (2015). Coverage of the Ebola virus disease epidemic on YouTube. *Disaster medicine and public health preparedness*, *9*(5), 531-535. https://doi.org/10.1017/dmp.2015.77

- Baumer, E. P. (2017). Toward human-centered algorithm design. *Big Data & Society*, *4*(2). https://doi.org/10.1177/2053951717718854
- Beer, D. (2017). The social power of algorithms. *Information Communication and Society*, *20*(1), 1–13. https://doi.org/10.1080/1369118X.2016.1216147

- Berg, M. (2014). Participatory trouble: Towards an understanding of algorithmic structures on Facebook. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *8*(3). https://doi.org/10.5817/CP2014-3-2
- Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-based systems*, *46*, 109-132. https://doi.org/10.1016/j.knosys.2013.03.012
- boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, *15*(5), 662-679. https://doi.org/10.1080/1369118X.2012.678878
- Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, *15*(3), 209–227. https://doi.org/10.1007/s10676-013-9321-6
- Bozdag, E., & Van Den Hoven, J. (2015). Breaking the filter bubble: democracy and design. *Ethics and information technology*, *17*(4), 249-265. https://doi.org/10.1007/s10676-015-9380-y
- Brunton, S. L., & Kutz, J. N. (2019). *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.
- Bryman, A. (2016). *Social research methods*. Oxford university press. http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf
- Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New media & society*, *14*(7), 1164-1180. https://doi.org/10.1177/1461444812440159

- Bucher, T. (2017). The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, communication & society*, *20*(1), 30-44. https://doi.org/10.1080/1369118X.2016.1154086

- Buckingham, D. (2007). Digital Media Literacies: rethinking media education in the age of the Internet. *Research in comparative and international education*, *2*(1), 43-55. https://doi.org/10.2304/rcie.2007.2.1.43

- Chaffey, D. (2016). Global social media research summary 2016. *Smart Insights: Social Media Marketing*.
- Chen, Y., Wu, C., Xie, M., & Guo, X. (2011). Solving the sparsity problem in recommender systems using association retrieval. *Journal of computers*, *6*(9), 1896-1902. https://doi.org/10.4304/jcp.6.9.1896-1902

- Covington, P., Adams, J., & Sargin, E. (2016). Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems*, 191-198. https://doi.org/10.1145/2959100.2959190

- Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., ... & Sampath, D. (2010). The YouTube video recommendation system.

  In *Proceedings of the fourth ACM conference on Recommender systems*, 293-296. https://doi.org/10.1145/1864708.1864770
- DeVito, M. A., Birnholtz, J., Hancock, J. T., French, M., & Liu, S. (2018). How people form folk theories of social media feeds and what it means for how we study self-presentation. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1-12. https://doi.org/10.1145/3173574.3173694
- Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, *59*(2), 56-62. https://doi.org/10.1080/21670811.2016.1208053

- Diakopoulos, N., & Koliska, M. (2017) Algorithmic Transparency in the News Media. *Digital Journalism,* 5(7), 809 - 828. https://doi.org/10.1080/21670811.2016.1208053

- Erviti, M. C., & Stengler, E. (2016). Online science videos: An exploratory study with major professional content providers in the United Kingdom. *Journal of Science Communication, 15*(6). https://doi.org/10.22323/2.15060206

- Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., ... & Sandvig, C. (2015). " I always assumed that I wasn't really that close to [her]" Reasoning about Invisible Algorithms in News Feeds. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 153-162. https://doi.org/10.1145/2702123.2702556

- Fast, N. J., & Jago, A. S. (2020). Privacy matters… or does It? Algorithms, rationalization, and the erosion of concern for privacy. *Current Opinion in Psychology*, *31*(1890), 44–48. https://doi.org/10.1016/j.copsyc.2019.07.011

- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, *80*(S1), 298-320. https://doi.org/10.1093/poq/nfw006

- Flick, U. (2009). An Introduction To Qualitative Fourth Edition. *SAGE Publications*, 506.
- French, M., & Hancock, J. (2017). What's the Folk Theory? Reasoning About Cyber-Social Systems. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2910571

- Fuchs, C. (2012). The political economy of privacy on Facebook. *Television & New Media*, *13*(2), 139-159. https://doi.org/10.1177/1527476411415699
- Garfinkel, S., Matthews, J., Shapiro, S. S., & Smith, J. M. (2017). Toward algorithmic transparency and accountability. In *Communications of the ACM, 60*(9), 5. https://doi.org/10.1145/3125780

- Gelman, S. A., & Legare, C. H. (2011). Concepts and folk theories. *Annual review of anthropology*, *40*, 379-398. https://doi.org/10.1146/annurev-anthro-081309-145822

- Gillespie, T. (2014). The relevance of algorithms. *Media technologies: Essays on communication, materiality, and society*, *167*(2014), 167. https://doi.org/10.7551/mitpress/9780262525374.003.0009

- Gillespie, T. (2016). trendingistrending: When algorithms become culture. *Algorithmic Cultures: Essays on Meaning, Performance and New Technologies, 189*(2016), 52–75. https://doi.org/10.4324/9781315658698
- Goldenfein, J. (2019). Algorithmic transparency and decision-making accountability: Thoughts for buying machine learning algorithms in Office of the Victorian Information Commissioner (ed). *Closer to the Machine: Technical, Social, and Legal aspects of AI.* https://doi.org/10.2861/536131
- Goldfarb, A., & Tucker, C. (2012). Shifts in privacy concerns. *American Economic Review*, *102*(3), 349-53. https://doi.org/10.1257/aer.102.3.349

- Hallinan, B., & Striphas, T. (2016). Recommended for you: The Netflix Prize and the production of algorithmic culture. *New media & society*, *18*(1), 117-137. https://doi.org/10.1177/1461444814538646

- Hindman, M. (2008). *The myth of digital democracy*. Princeton University Press. https://doi.org/10.5860/choice.46-5888
- Introna, L. D. (2011). The enframing of code: Agency, originality and the plagiarist. *Theory, Culture & Society*, *28*(6), 113-141. https://doi.org/10.1177/0263276411418131

- Iphofen, R., Tolich, M., Delamont, S., & Atkinson, P. (2019). The Ethics of Ethnography. *The SAGE Handbook of Qualitative Research Ethics*, 119–132. https://doi.org/10.4135/9781526435446.n8

- Isaias, P., Miranda, P., & Pifano, S. (2013). Digitally literate: Framework for the development of digital literacy. *Proceedings of the IADIS International Conference WWW/Internet 2013, ICWI 2013*, *1997*, 277–282.

- Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, *16*(3), 261-273. https://doi.org/10.1016/j.eij.2015.06.005

- Jespersen, J. L., Albrechtslund, A., Øhrstrøm, P., Hasle, P., & Albretsen, J. (2007). Surveillance, persuasion, and panopticon. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *4744*, 109–120. https://doi.org/10.1007/978-3-540-77006-0_15

- Kim, J. (2012). The institutionalization of YouTube: From user-generated content to professionally generated content. *Media, culture & society*, *34*(1), 53-67. https://doi.org/10.1177/0163443711427199

- Koene, A., Clifton, C., Hatada, Y., Webb, H., & Richardson, R. (2019). *A governance framework for algorithmic accountability and transparency*. Brussels: European Parliamentary Research Service. https://op.europa.eu/en/publication-detail/-/publication/8ed84cfe-8e62-11e9-9369-01aa75ed71a1/language-en

- Lankshear, C., & Knobel, M. (2015). Digital Literacy and Digital Literacies: Policy, Pedagogy and Research Considerations for Education. *Nordic Journal of Digital Literacy*, *9*(4), 8–20.

- Lee, M. K., Kim, J. T., & Lizarondo, L. (2017). A human-centered approach to algorithmic services: Considerations for fair and motivating smart community service management that allocates donations to non-profit organizations.

In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3365-3376. https://doi.org/10.1145/3025453.3025884

- Leonardi, P. M., Nardi, B. A., & Kallinikos, J. (Eds.). (2012). *Materiality and organizing: Social interaction in a technological world*. Oxford university press. https://doi.org/10.1093/acprof:oso/9780199664054.001.0001

- Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology*, *31*(4), 611-627. https://doi.org/10.1007/s13347-017-0279-x

- Lewis, R. (2018). Alternative influence: Broadcasting the reactionary right on YouTube. *Data & Society*, *18*.
- Li, H., Cai, F., & Liao, Z. (2012). Content-based filtering recommendation algorithm using HMM. In *2012 Fourth International Conference on Computational and Information Sciences* , 275-277. IEEE. https://doi.org/10.1109/ICCIS.2012.112
- Li, X., Xing, J., Wang, H., Zheng, L., Jia, S., & Wang, Q. (2018). A Hybrid Recommendation Method Based on Feature for Offline Book Personalization. *Journal of Computers, 30*(5), 1-17. https://doi.org/10.3966/199115992019103005001

- Linden, G., Smith, B., & York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, *7*(1), 76-80. https://doi.org/10.1109/MIC.2003.1167344

- Livingstone, S. (2003). The changing nature and uses of media literacy.
- Loisen, J., & Joye, S. (2017). *On media & communication : an introduction to*

communication sciences : theory and research. Acco.
https://lib.ugent.be/catalog/pug01:8540040

- Madra-Sawicka, M., Nord, J. H., Paliszkiewicz, J., & Lee, T. R. (2020). Digital media: Empowerment and equality. *Information (Switzerland)*, *11*(4). https://doi.org/10.3390/INFO11040225

- Malle, B. F. (2009). Folk theories of consciousness. *Encyclopedia of consciousness*, *1*, 251-263. https://doi.org/10.1016/B978-012373873-8.00023-2

- Mason, J. (2002). Qualitative Researching 2nd Edition. In *Qualitative Research Book*.

- Mitchell, A., Rosenstiel, T., & Christian, L. (2012). *The state of the media 2012: What Facebook and Twitter mean for news*. *Media*, 1–9.
- Munger, K., & Phillips, J. (2019). A Supply and Demand Framework for YouTube Politics. *OSF Preprint*.
- Munoz, C., Smith, M., & Patil, D. (2016). Big Data : A Report on Algorithmic Systems, Opportunity, and Civil Rights. *Executive Office of the President of USA*. Retrieved from https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

- Napoli, P. M. (2014). Automated media: An institutional theory perspective on algorithmic media production and consumption. *Communication Theory*, *24*(3), 340–360. https://doi.org/10.1111/comt.12039

- Napoli, P. M. (2015). Social media and the public interest: Governance of news platforms in the realm of individual and algorithmic gatekeepers. *Telecommunications Policy*, *39*(9), 751–760. https://doi.org/10.1016/j.telpol.2014.12.003

- Oh, C., Lee, T., Kim, Y., Park, S., Kwon, S., & Suh, B. (2017). Us vs. Them: Understanding artificial intelligence technophobia over the google deepmind challenge match. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2523-2534. https://doi.org/10.1145/3025453.3025539

- Pasquale, F. (2015). *The black box society*. Harvard University Press. https://doi.org/10.4159/harvard.9780674736061
- Rader, E., & Gray, R. (2015). Understanding user beliefs about algorithmic curation in the Facebook news feed. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 173-182. https://doi.org/10.1145/2702123.2702174
- Rader, E., Cotter, K., & Cho, J. (2018). Explanations as mechanisms for supporting algorithmic transparency. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1-13. https://doi.org/10.1145/3173574.3173677
- Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, *40*(3), 56-58. https://doi.org/10.1145/245108.245121
- Saputra, M., & Al Siddiq, I. H. (2020). Social media and digital citizenship: The urgency of digital literacy in the middle of a disrupted society Era. *International Journal of Emerging Technologies in Learning*, *15*(7), 156–161. https://doi.org/10.3991/IJET.V15I07.13239
- Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In *The adaptive web*, 291-324. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-72079-9_9
- Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, 253-260.

- Siersdorfer, S., Chelaru, S., Nejdl, W., & San Pedro, J. (2010). How useful are your comments? Analyzing and predicting YouTube comments and comment ratings. In *Proceedings of the 19th international conference on World wide web*, 891-900. https://doi.org/10.1145/1772690.1772781

- Simon, B. (2005). The return of Panopticism: Supervision, subjection and the new surveillance. *Surveillance & Society*, *3*(1). https://doi.org/10.24908/ss.v3i1.3317

- Strauss, A., & Corbin, J. (1990). *Basics of Qualitative research: Grounded theory procedures and techniques*. Newbury.
- Taylor, C. R. (2004). Consumer privacy and the market for customer information. *RAND Journal of Economics*, 35(4), 631-650. https://doi.org/10.2307/1593765

- The European Commission. (2020). *Register of Commission Expert Groups.* Retrieved from https://ec.europa.eu/transparency/regexpert/index.cfm?do=search.resultNew

- The European Parliament and the Council of the European Union. (2016*). Regulation EU 2016/679 of the European Parliament and of the Council of 27 April 2016*. Brussels: Official Journal of the European Union. http://ec. europa. eu/justice/data-protection/reform/files/regulation_oj_en. pdf

- The European Parliament and the Council of the European Union. (2007). *Directive 2007/65/EC of the European Parliament and of the Council of 11 December 2007 amending Council Directive 89/552/EEC on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the pursuit of television broadcasting activities*.

Strasbourg.

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32007L0065

- Toffler, A. (1980). *The third wave*. New York: Bantam books.
- Tucker, C. (2018). Privacy, algorithms, and artificial intelligence. *The Economics of Artificial Intelligence: An Agenda*, 423–437. https://www.nber.org/system/files/chapters/c14011/c14011.pdf
- Tufekci, Z. (2015). Facebook said its algorithms do help form echo chambers, and the tech press missed it. *New Perspectives Quarterly*, *32*(3), 9-12. https://doi.org/10.1111/npqu.11519
- Van Deursen, A. J., & van Dijk, J. A. (2009). Improving digital skills for the use of online public information and services. *Government Information Quarterly*, *26*(2), 333-340.
- Vozalis, E., & Margaritis, K. G. (2003). Analysis of recommender systems algorithms. In *The 6th Hellenic European Conference on Computer Mathematics & its Applications*, 732-745.
- Welbourne, D. J., & Grant, W. J. (2016). Science communication on YouTube: Factors that affect channel and video popularity. *Public understanding of science*, *25*(6), 706-718. https://doi.org/10.1177/0963662515572068
- Willson, M. (2017). Algorithms (and the) everyday. *Information, Communication & Society*, *20*(1), 137-150. https://doi.org/10.4324/9781351200677-9
- Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values*, *41*(1), 118-132. https://doi.org/10.1177/0162243915605575

- Zhang, T., & Iyengar, V. S. (2002). Recommender systems using linear classifiers. *The Journal of Machine Learning Research*, *2*, 313-334. https://doi.org/10.1162/153244302760200641

- Zhao, Z., Hong, L., Wei, L., Chen, J., Nath, A., Andrews, S., ... & Chi, E. (2019). Recommending what video to watch next: a multitask ranking system. In *Proceedings of the 13th ACM Conference on Recommender Systems*, 43-51. https://doi.org/10.1145/3298689.3346997

- Zhou, R., Khemmarat, S., & Gao, L. (2010). The impact of YouTube recommendation system on video views. In *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, 404-410. https://doi.org/10.1145/1879141.1879193
- Zuboff, S. (2019a). Surveillance capitalism and the challenge of collective action. In *New labor forum*, *28*(1), 10-29. SAGE Publications. https://doi.org/10.1177/1095796018819461

- Zuboff, S. (2019b). *The age of surveillance capitalism: The fight for a human future at the new frontier of power: Barack Obama's books of 2019*. Profile books.

NON-ACADEMIC LITERATURE

- Amazon Web Services. (2020). *Alexa Top Sites*. Amazon Web Services. https://www.alexa.com/topsites
- Antonio, A., & Tuffley, D. (2015, January 15). *YouTube a valuable educational tool, not just cat videos.* The Conversation. https://theconversation.com/youtube-a-valuable-educational-tool-not-just-cat-videos-34863
- Aslam, S. (2021, January 6). *YouTube by the Numbers: Stats, Demographics and Fun Facts.* Omnicore. https://www.omnicoreagency.com/youtube-statistics/

- Creator Insider. (2017, August 18). *YouTube Algorithm: YouTube employees answer your questions! (part 1).* https://youtu.be/QwdqXpe1R7c
- Federal Trade Commission. (2019, September 4). *Google and YouTube Will Pay Record $170 Million for Alleged Violations of Children's Privacy Law*. FTC. https://www.ftc.gov/news-events/press-releases/2019/09/google-youtube-will-pay-record-170-million-alleged-violations
- Hale, J. (2019, May 7). *More Than 500 Hours Of Content Are Now Being Uploaded To YouTube Every Minute.* Tubefilter. https://www.tubefilter.com/2019/05/07/number-hours-video-uploaded-to-youtube-per-minute/
- Jeanne Dugan, I. (1996, October 21). *`The Internet Is The Great Equalizer'*. Bloomberg. https://www.bloomberg.com/news/articles/1996-10-20/the-internet-is-the-great-equalizer
- Kanter, J. (2018, May 20). *The backlash that never happened: New data shows people actually increased their Facebook usage after the Cambridge Analytica scandal*. Business Insider. https://www.businessinsider.nl/people-increased-facebook-usage-after-cambridge-analytica-scandal-2018-5/?international=true&r=US&jwsource=cl
- Leonhardt, M. (2019, July 22). *What you need to know about the Equifax data breach $700 million settlement.* CNBC. https://www.cnbc.com/2019/07/22/what-you-need-to-know-equifax-data-breach-700-million-settlement.html
- Nielsen, J. (1994, April 24). *10 Usability Heuristics for User Interface Design.* Nielsen Norman Group. Retrieved April 13, 2021 from https://www.nngroup.com/articles/ten-usability-heuristics/

- Nunez, M. (2016, May 16). *Former Facebook Workers: We Routinely Suppressed Conservative News*. Gizmodo. https://gizmodo.com/former-facebook-workers-we-routinely-suppressed-conser-1775461006

- Pandey, A. (2020, August 5). *How to improve recommendations for highly sparse datasets using Hybrid Recommender Systems?* Artificial Intelligence. https://ai.plainenglish.io/how-to-improve-recommendations-for-highly-sparse-datasets-using-hybrid-recommender-systems-1a4366e65cff

- Roose, K. (2019, March 29). *YouTube's product chief on online radicalization and algorithmic rabbit holes.* The New York Times. https://www.nytimes.com/2019/03/29/technology/youtube-online-extremism.html

- Solsman, J. E. (2018, January 10). *YouTube's AI is the puppet master over most of what you watch.* CNET. https://www.cnet.com/news/youtube-ces-2018-neal-mohan/

- VPRO Documentary. (2019, 21 December). *Shoshana Zuboff on surveillance capitalism.* https://youtu.be/hIXhnWUmMvw
- Wagner, A. (2019, May 15). *Are You Maximizing The Use Of Video In Your Content Marketing Strategy?* Forbes. https://www.forbes.com/sites/forbesagencycouncil/2017/05/15/are-you-maximizing-the-use-of-video-in-your-content-marketing-strategy/
- Warzel, C. (2017, October 4). *Here's How YouTube Is Spreading Conspiracy Theories About The Vegas Shooting*. BuzzFeed News. https://www.buzzfeednews.com/article/charliewarzel/heres-how-youtube-is-spreading-conspiracy-theories-about

- YouTube Creator Academy. (n.d.). *Search and discovery on YouTube - YouTube*. 2019. Retrieved March 22, 2021, from https://creatoracademy.youtube.com/page/lesson/discovery