Algorithmic Accountability and the Right to Information:
Visibility & Invisibility of Online Content During Elections

LLM Paper
by Anne Oloo
Student number: 01702934
Promoter: Prof. Dr. Eva Lievens
ACKNOWLEDGEMENTS

I wish to thank first and foremost my promoter Prof. Dr. Eva Lievens, without whom the completion of this paper would not have been possible. I am extremely grateful for her enthusiasm with the topic of this paper, her insightful suggestions and comments, her patience, understanding and support throughout the process of planning and writing this paper. I would also like to acknowledge Ms. Ingrida Milkaite, Doctoral Student at Ghent University, as the second reader of this paper. I benefitted greatly from her valuable comments and suggestions.

There are several people whom I met at Ghent university that I wish to acknowledge. First, to my Jessup Teammates and coaches Kiana, Marlies, Michiel, Andrés, Joyce and Laurence and Professor Frank Maes, thank you for making this past year the most exciting and rewarding year yet. Not only did I gain valuable research, analytic and teamwork skills but I also made friends who were supportive of my other endeavours such as the writing of this paper.

I wish to thank my friends and fellow colleagues from the LL.M Programme. A special mention to Devanshi Saxena and Suay Boulougouris for their friendship and constant support throughout the LL.M programme.

I am also grateful to Dr. Luis Franceschi (Dean, Strathmore University) and Prof. Edwin Abuya (University of Nairobi) who supported me from the moment I expressed interest to further my studies in law.

Lastly, I acknowledge the special place of my family, who have been a constant source of support and encouragement. My deepest gratitude to my father Washington Oloo and my mother, Catherine Odera, to all my siblings and my parents-in-law for always believing in me and for their love and support in everything I have pursued. Most importantly, I am eternally grateful for my husband, Damien, whose love, emotional and moral support and understanding made it all possible.
# Table of Contents

Acknowledgements .......................................................................................................................... ii

Chapter 1: Introduction ...................................................................................................................... 1
  1.1. Introduction and statement of objectives ............................................................................. 1
  1.2. Scope of study and limitations ............................................................................................. 2

Chapter 2: Algorithms, machine learning, tailored news: a brief introduction ............................ 3
  2.1. Algorithms ............................................................................................................................. 3
  2.2. Introduction to the Facebook Algorithm ............................................................................. 4

Chapter 3: The right to information ................................................................................................. 9
  3.1. Introduction ............................................................................................................................. 9
  3.2. Scope of RTI .......................................................................................................................... 11
  3.3. RTI during elections .............................................................................................................. 16
  3.4. Conclusion ............................................................................................................................. 25

Chapter 4: Algorithmic accountability .......................................................................................... 27
  4.1. Introduction ............................................................................................................................. 27
  4.2. The problem with algorithms and online platforms .............................................................. 29
  4.3. Algorithmic accountability: Definition and principles ......................................................... 36
  4.4. Methods to achieve algorithmic accountability ................................................................. 45

Chapter 5: Right to information, algorithmic accountability, visibility ......................................... 51
  5.1. Right to information and the (in)visibility of online content ............................................. 51
  5.2. Fostering algorithmic accountability to ensuring a practical and effective right to information on online platforms ................................................................. 55
  5.3. Protecting freedom of speech and the right to information worldwide ......................... 60

BIBLIOGRAPHY ............................................................................................................................. 63
CHAPTER 1: INTRODUCTION

1.1. Introduction and statement of objectives

The internet has revolutionised the way in which we receive and impart information. In the age of Big Data and social media platforms, many receive their news from the tailored news feeds provided by Facebook and similar platforms. However, the algorithms governing the selection of items (including news stories) appearing in individual feeds are often opaque, closely guarded as trade secrets by companies. This implies that users have little power over and awareness of the hidden selection mechanisms that determine what news they will and will not see. This raises concerns about the compatibility of such mechanisms with the right to information, a key principle of democratic societies. Recent events, such as the use of Facebook data by firms such as Cambridge Analytica in elections,\(^1\) further underline the challenges surrounding the ubiquity of social media platforms. The right to seek, receive and impart information takes an even greater importance during elections, underpinning citizens’ ability to make an informed decision about the candidates and the issues. Tailored content thus raises significant questions about the role of online platforms in elections and their compatibility with the need to having a well-informed citizenry capable of comparing what all candidates propose on the issues of the day.

As a reaction, calls for ‘algorithmic accountability’ are increasingly made. In this perspective, the all-powerful, hegemonic platforms of today should better serve the interests of their users by becoming more transparent and ensuring that no algorithmic harms are committed. In the context of elections, this would require that online platforms do not, through their action or inaction, distort the electoral process. Yet, while such calls identify principles to be safeguarded, they do not necessarily connect them to broader rights protected by international instruments. Such an approach is however beneficial, as it provides firmer footing to regulatory measures and ensures that proposed regulation does not itself run afoul of human rights. This paper will thus examine visibility of online content during elections under the prism of the right to information (RTI).

The present paper will first introduce the concept of algorithms and the general principles governing the visibility and invisibility of information on Facebook. It will then provide an overview of RTI and

of algorithmic accountability, before bringing these two concepts to bear on the question of personalised content during elections. The implications of RTI for the tailoring of online content will be drawn and combined with algorithmic accountability to create a framework to protect and foster RTI on online platforms during elections. Policy proposals will finally be put forward and assessed in light of the different aspects of RTI.

1.2. Scope of study and limitations

This paper is necessarily limited in scope. The topic of algorithmic accountability has been widely discussed and online platforms, from Twitter, to Youtube, to Facebook, all have their own idiosyncrasies. This makes it necessary to limit the scope of the paper which will focus on online content, especially news, using Facebook as a case study whenever it is necessary to analyse a particular platform. This is because Facebook is the most prominent social media platform and has been central to the rise of awareness about algorithmic accountability due to the Cambridge Analytica scandal.

To explore media regulation during elections, two countries were chosen. First, France, which was chosen because it is a European country with an extensive media regulation framework. It has produced several reports on algorithmic accountability and has recently held elections during which social media played a role. Second, Kenya, which also recently held contested elections influenced by social media. Including a non-Western country with a less favourable media landscape allows to take into account the need to protect individuals against both private and state interference with RTI.
CHAPTER 2: ALGORITHMS, MACHINE LEARNING, TAILORED NEWS: A BRIEF INTRODUCTION

2.1. Algorithms

Much of the literature and policy debate centres around the role of ‘algorithms’ which can be defined in the broadest sense as ‘methods for solving problems that are suited for computer implementation’\(^2\). Algorithms need not be complicated or innovative: for instance, both Euclid’s method for finding the greatest common divisor of two numbers (first formulated around 300 BCE)\(^3\) and a method to efficiently sort a list\(^4\) are algorithms.

Algorithms are thus ubiquitous in our modern world, with most of their uses uncontroversial and mundane. Elevators use an algorithm to determine which floors to serve first.\(^5\) Air conditioning relies on an algorithm to efficiently bring the room to the desired temperature. File compression programs implement algorithms that shrink files while retaining all the (important) information.

The debate surrounding algorithms stems from their use in domains where ethical considerations are primordial (e.g. medicine, the legal process, human resources) or which have rapidly become central to our lives, with little understanding of their impact. Chief among those is the rise of social medial platforms, which tailor displayed contents to the (perceived) interests of their users while being attuned to the needs of their business partners, as they rely on ads and other partnerships to monetise their platforms. In this context, algorithms are used to identify which contents to display first, or at all, in order to maximise user satisfaction, profitability, user engagement, or any other metric.

The mystique surrounding algorithms often makes them appear more mysterious than they are. Concepts such as ‘machine learning’, ‘supervised learning’, ‘unsupervised learning’ and ‘deep learning’ are used without always being understood. Although they can be explained in simple terms, this paper will not delve into technical details. It concerns itself not with the precise techniques utilised by online platforms but with the legal and ethical questions raised by the use. This does not require an

---

\(^3\) ibid 4.
\(^4\) ibid 243.
\(^5\) See e.g. DL Pepyne and CG Cassandras, ‘Optimal Dispatching Control for Elevator Systems during Uppeak Traffic’ (1997) 5 IEEE Transactions on Control Systems Technology 629, 629–630. Ever increasing building heights made it necessary to adopt semi-automatic dispatching procedures as elevator attendants could no longer cope with the complexity of the task. Note that even this seemingly uncontroversial use of algorithms could still prove controversial in some cases, for instance if a hotel prioritised luxury floors and inflicted longer waiting times on other unsuspecting guests.
explanation of the machine learning techniques, but rather of the way content is filtered and ranked by online platforms. The next subsection will thus introduce the Facebook ranking algorithm.

2.2. Introduction to the Facebook Algorithm

This subsection will briefly explain how Facebook filters and orders posts, both for organic content (content shared by users) and for sponsored posts. As these techniques are proprietary, only limited information is publicly available.

2.2.1. Organic posts

Whenever a user accesses their feed, Facebook is faced with a conceptually simple problem: deciding which posts should be included in this feed and in what order. To do so, it must select the most relevant posts in a potentially very large pool of posts vying for limited screen space. Pew reported in 2013 that teen Facebook users had a median number of friends of 300. As Facebook can display any action taken by each of these friends (posting, liking a post, writing a comment) or by third parties connected to these friends (e.g. tagging them in a photo), the number of posts to choose from can easily number in the thousands, while users will probably not scroll down further than a few dozen posts. This makes filtering and ordering a necessity to preserve the usability of the feed.

In its earlier days, Facebook ordered its news feed based on three factors: affinity (how close the two users were, as measured by the frequency of interaction), weight (what importance should be given to each type of content, and time decay (older content is less important than newer content). Since 2013, Facebook has used a machine learning algorithm. Instead of assigning weights to different types of content and developing metrics to measure affinity, Facebook surveyed many users to collect their reaction to different types of posts. This data was then used to predict which posts would be considered most relevant.

---

10 Kacholia (n 8).
Since then, Facebook has constantly tweaked its algorithm\(^{11}\), most recently reacting to growing backlash by announcing that it would prioritise local, trustworthy news and meaningful interactions between individuals. The list of changes however also testifies to the pervasiveness of data collection that characterises Facebook, as each ‘improvement’ relies on further surveillance of for instance time spent reading articles to which posts link, or of how much of a video a user watches.

Recently, information shows that Facebook has fully embraced the machine learning paradigm and that most decisions are based on user supplied data, even when Facebook adopts a specific prioritisation policy. Thus, Facebook relies on user surveys to determine which news sources are trustworthy\(^{12}\) and uses past interactions to predict which type of content is most likely to spark meaningful conversations\(^{13}\).

Despite limited efforts to better explain the philosophy of its news feed, Facebook has so far revealed little about the inner workings of its algorithm. Explanations have focused on broad principles\(^{14}\) that provide little information about crucial steps in the ranking process. For instance, how the score used to order stories is computed is not explained. No information is given on how different actions are valued by the platform (e.g. is a comment ‘worth’ more than a ‘like’ or a ‘share’? Are commercial factors, e.g. increased exposition for partner brands, taken into account?). Facebook primarily informs users that it possesses large amounts of information about them and use them in unspecified ways to determine which content is most relevant. Even deeper discussions of the algorithm typically only reveal general principles, e.g. that the relevancy score of unreliable news is multiplied ‘by some fraction’\(^{15}\).

Finally, no information is available regarding the ranking of individual posts, nor is it possible to manually tweak the order in which posts are shown. Users can only influence the ranking of future posts by reacting to current ones, for instance by muting specific friends, hiding posts to signal lack of interest in such content, or by liking, sharing, or commenting on posts, which should increase the rank

---


of similar posts in the future (with unknown magnitude). Some third-party effort toward this goal have however been done, for instance through the FeedVis application\textsuperscript{16}, which allowed users to explore hidden stories and compare their news feed to what they might have been shown. This allowed users to identify friends that were underrepresented in their feed. One limitation is that this was done in the context of a research project and cannot be scaled since applications cannot recreate an individual’s feed without permission from the friends that generate the content.

2.2.2. Sponsored posts

Sponsored posts are ads shown alongside organic content in a seamless manner. Facebook presents such content using the same layout as organic content, distinguishing the two solely through a small ‘sponsored’ mention that replaces the date of posting.

All users can sponsor posts, which provides valuable information about the inner workings of sponsored posts. This is the method that was used to write these paragraphs. Facebook allows advertisers to choose a marketing objective e.g. raising brand awareness, creating video views, lead generation or store visits, and a target audience. This target audience can be highly tailored using the troves of data that Facebook holds. Besides traditional demographic variables such as age or gender, more specific categories are available. Advertisers can thus choose to target users ‘likely to engage with political content (conservative)’ or users with various ‘multicultural affinities’ (e.g. African American, LGBTs). While Facebook prohibits discriminatory practices\textsuperscript{17} and restricts the categories that can be used to exclude people, there seems to be little systematicity and consistency, leaving open the possibility of discrimination through exclusion of overlooked categories that correlate with protected characteristics (e.g. excluding those who have liked Essence magazine to exclude many African-American women).

Advertisers can also make use of their own data to create custom audiences. Websites can include a hidden Facebook pixel that tracks visitors and shows ads to users who have already visited the website. Advertisers can also target people who have already interacted with their business outside Facebook, for instance by visiting their store.


\textsuperscript{17} ‘Advertising Policies’ (Facebook) <https://www.facebook.com/policies/ads/prohibited_content/discriminatory_practices> accessed 23 February 2018.
Once a campaign is created, Facebook optimises how sponsored posts are shown to achieve the predetermined goal (e.g. increase brand awareness) while remaining within the budget limits set by the advertiser.

In contrast with organic posts, Facebook displays some information regarding how sponsored posts are chosen. Users can obtain basic information about the criteria that determined that a sponsored post should be shown, although this information is often very limited (e.g. only mentioning that a post was shown because the user speaks English and is over 18, while it is likely that many other factors, such as previous visits on the advertiser’s website or friends’ engagement with the same advertiser, were taken into account). Thus, while Facebook gives the semblance of transparency necessary for users to retain confidence in the platform, there is little genuine information.
CHAPTER 3: THE RIGHT TO INFORMATION

3.1. Introduction

The term ‘right to information’ (RTI) has two different, but related, meanings:

1. First, ‘right to information’ is often used as a synonym of ‘freedom of information’, to refer to the right to obtain information, primarily documents, held by state organs. This is evidenced by the growing number of ‘Access to information Acts’ enacted in various jurisdictions.\(^{18}\)
2. Second, ‘right to information’ also refers, broadly, to the right to inform and to be informed, with all accompanying requirements and consequences such as freedom of the press\(^{19}\) and freedom of speech and of communication\(^{20}\).

These two aspects are closely related: both find their origin in the belief that an informed citizenry is a prerequisite of a well-functioning democratic society. To effectively exercise their rights, citizens need access to relevant information, as recognised for instance by Article 1 of the Aarhus Convention\(^ {21}\) in the context of environmental information or General Comment 34 with regards to elections\(^ {22}\).

The connexion between the two aspects of RTI can also be seen in a historical perspective: the first Freedom of Information Act, granting access to many administrative and legal documents, was Sweden’s Ordinance Relating to Freedom of Writing and of the Press (1766)\(^ {23}\). The inclusion of this right to information in a text protecting freedom of writing and of the press highlights the tight link between the right to obtain official documents and the ability of the press and of authors to publish information gleaned from these documents, in order to enable citizens to receive this information.

---


\(^{20}\) Pekka Hallberg and Janne Virkkunen, Freedom of Speech and Information in Global Perspective (Springer 2017) 81–82.


RTI has since been recognised in numerous international and regional instruments\(^\text{24}\) and is firmly enshrined in international human rights law. It is provided for in the UDHR\(^\text{25}\), which constitutes customary international law\(^\text{26}\), the ICCPR\(^\text{27}\) and ICESCR\(^\text{28}\) under Article 19 of each of these instruments, which state (with slightly variant formulations) that

> Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and **to seek, receive and impart** information and ideas through any media and regardless of frontiers. (UDHR wording, emphasis added)

Likewise, Article 10(1) of the ECHR provides that

> Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and **to receive and impart** information and ideas without interference by public authority and regardless of frontiers. This article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises.\(^\text{29}\)

Comparing the wording of the two instruments shows that, while the two aspects of RTI are often linked, they are not necessarily explicitly so. The UDHR thus explicitly protects the freedom to seek information, which underpins freedom of information laws. Consequently, UN bodies have long stressed the importance of access to publicly-held information: for instance, the United Nations General Assembly (UNGA) has proclaimed that ‘freedom of information is a fundamental human right and the touchstone for all freedoms to which the United Nations is consecrated’\(^\text{30}\), a sentiment later echoed by Special Rapporteur Abdi Hussain\(^\text{31}\). By contrast, the ECtHR had long been reluctant to


\(\text{25} \) Universal Declaration of Human Rights (adopted 10 December 1948) UNGA Res 217 A (III) (UDHR).

\(\text{26} \) Anna Södersten, *A Companion to European Union Law and International Law* (John Wiley & Sons 2016) 444.

\(\text{27} \) International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 (ICCPR).


\(\text{30} \) ‘UNGA Res 59 (1) Calling of an International Conference on Freedom of Information (14 December 1946)’.

recognise a right to seek information not explicitly included in the ECHR. However, recent jurisprudence shows a move towards acceptance that freedom to receive information includes access to information.

It is primarily the second aspect of the RTI that will be further developed in this chapter. Indeed, while the first aspect is undoubtedly important to the rule of law, it has less direct bearing on the issue of the visibility and invisibility of online news during elections. This chapter will thus focus on the scope of the second meaning of RTI (the right to inform and to be informed, excluding access to official documents). It will first provide an overview of the scope and contents of RTI. It will then identify the implications of RTI during elections. An overview of media regulation, with particular focus on election periods, will finally be provided in order to identify ways in which RTI is both protected and restricted on such occasions.

3.2. Scope of RTI

While RTI in human right treaties is usually included in the right to freedom of expression, it has also been recognised as a separate right consisting of the right to seek, receive and impart information. The right to seek information is associated with the first aspect of RTI described above i.e. all that concerns access to publicly-held information held by the state and ‘the corresponding duty upon the state to make that information available’, while the second aspect, the right to inform and to be informed, is mainly embodied in the right to receive and impart information. These last two freedoms are complementary and one cannot exist without the other. Indeed, Justice Brennan observed that:

‘the right to receive ideas follows ineluctably from the sender’s First Amendment right to send them: [...] The dissemination of ideas can accomplish nothing if otherwise willing addressees are not free to receive them. It would be a barren marketplace of ideas that had only sellers and not buyers.’

35 Andrew Puddephatt, ‘Right to Information: Practical Guidance Note’ 5
As noted above, the distinction is not water-tight as seeking information can constitute a preliminary step to receiving or imparting it.

3.2.1. RTI as a prohibition on (state) interference

Traditionally, the right to receive and impart information has applied vertically rather than horizontally\(^{38}\), i.e. it is enforced against the state rather than applied to relations between private parties.\(^{39}\) This reflects the historical approach of human rights systems, which were originally intended to protect individuals from abuses of state power. In this framework, the state has the responsibility to protect rights by not interfering with the free exercise of these rights.\(^{40}\) This means that human rights provisions are primarily addressed to the state, as these rights are seen to apply only in the public sphere. By contrast, the private sphere should be protected from state intrusion.\(^{41}\)

Applied to RTI, this view involves focus on the myriad of ways in which states can restrict the freedom to receive and impart information. Censorship has historically been an important factor limiting these freedoms. Thus, RTI applied vertically aims to protect the right to freely disseminate ideas and to freely receive them by prohibiting censorship and greatly limiting the ability of the state to interfere with the free flow of information. This prohibition has been recognised to apply to all forms of information including information that may ‘shock, offend or disturb\(^{42}\), information of no social worth\(^{43}\), commercial speech\(^{44}\) and political speech\(^{45}\); expressed through all forms of media\(^{46}\) (print media, broadcast media, the internet) and regardless of frontiers. Courts have thus held as violations of RTI various types of state interference: prohibiting a newspaper from publishing an article\(^{47}\), refusing the registration of a periodical\(^{48}\), preventing a union president from making a speech to recruit new members without a special permit\(^{49}\), preventing corporations from disseminating information about political issues\(^{50}\), or fining a political party for making available a smartphone application.

---


\(^{39}\) ibid.


\(^{41}\) Corrin (n 40).

\(^{42}\) Handyside v UK [1976] ECHR App No. 5493/72 [49].


\(^{44}\) See X and Church of Scientology v Sweden [1979] ECHR App No. 7805/77.


\(^{48}\) Gaweda v Poland [2002] ECHR (Former Section I) App No. 26229/95.


allowing voters to take pictures of their spoilt ballot. Likewise, restricting the right to receive information through blocking YouTube or all Google sites has also been held to be a violation of RTI. Special attention has been paid to freedom of the press and on state interference with it because of the important watchdog role that the media plays in democratic society and of their status as traditional purveyors of information of public interest. However, it has been stressed by the UN Human Rights Committee and the Council of Ministers of the Council of Europe that a functional definition of journalist should be adopted to incorporate new practices such as blogging: ‘journalism is an activity, which consists of the collection and dissemination of information to the public via any means of mass communication.

This does not mean that this protection is absolute, as RTI may conflict with other considerations (e.g. privacy, national security, rights of minorities). Thus, RTI may be restricted if that restriction is prescribed by law, pursues a legitimate aim and is necessary and proportionate in a democratic society. The degree of permissible restriction however varies between jurisdictions and legal traditions, for instance between most European countries and the United States.

3.2.2. RTI and positive obligations

The vertical approach to rights has been criticised by for failing to take into account the changing nature of the public sphere. Although private entities have increasingly been involved in the provision of public resources and services, a purely vertical approach leaves individuals without means to invoke their right to receive and impart information against private institutions (such as online platforms) that are considered as ‘de facto gatekeepers’ of information. Private individuals and companies can often limit one’s ability to receive or impart information as effectively as state authorities. For instance, the ECtHR found that being evicted by one’s landlord from a rented flat for

52 Cengiz v Turkey [2015] ECHR (Second Section) App No. 48226/10 and 14027/11.
56 ICCPR (n 27) art 19(3); ECHR (n 29) art 10(2).
59 Florini (n 58) 237.
60 Tufekci (n 9) 206.
installing a satellite dish was an interference with RTI. There is therefore an incipient tendency towards recognising the need for horizontal application of RTI, although this approach remains controversial and can differ from the original intention of the drafter of human rights instruments. This horizontal application can thus be by way of imposing positive obligations on the state to intervene to prevent the infringement of rights by third parties.

Such positive obligations can take two forms: having the right legal framework to prevent infringement by third parties, and taking appropriate measures to protect individuals when the exercise of RTI is threatened by third parties. One important objective of such positive obligations is to establish an enabling environment for the dissemination of information, for instance through the protection of journalists but also through measures aimed at safeguarding pluralism.

### 3.2.3. RTI and pluralism

Pluralism ‘essentially requires that as many different voices as possible shall be heard in the marketplace of ideas, such as different political, moral, cultural and religious opinions’. It must be promoted and safeguarded in order for RTI to be effective access to a wide variety of viewpoints is indeed necessary for democratic debate. This applies not only to opinions but also to news, as there is always a subjective element to which news are reported and to which type of coverage is given to an issue. Such decisions can have an important impact on citizens, as agenda-setting theory has established:

The repetition of messages about public issues in the news day after day, along with the pervasiveness of the mass media in our daily lives, constitute a major source of

---

61 Khurshid Mustafa and Tarzibachi v Sweden [2008] ECHR (Third Section) App No. 23883/06.
64 ibid 50.
66 Dink v Turkey [2010] ECHR (Second Section) App No. 2669/07 [137].
journalism’s influence on the audience. The incidental nature of this learning, in turn, helps issues to move rather quickly from the media agenda to the public agenda.\textsuperscript{69}

Thus, decisions by the media regarding which issues to cover, with what frequency and in which manner, influence what the public thinks about and how to think about these topics.\textsuperscript{70}

Pluralism guarantees that the public will be exposed to a wide variety of news and viewpoints. It requires the diversity of media supply, use and distribution, through diversity in ownership and control of the media, in media types and genres, in political viewpoints, in cultural expressions and in terms of local and regional interests.\textsuperscript{71} To foster this diversity, the state should both refrain from unduly restricting media pluralism (e.g. by imposing a state monopoly on broadcasting\textsuperscript{72}) and take positive action to create the conditions that will allow media pluralism to emerge. A pluralistic media landscape is one where there is variety of types of media and diverse media ownership, excluding both public and private monopoly over the media.\textsuperscript{73} As noted by Voyenne: ‘tout monopole est incompatible avec l’information et un monopole d’Etat l’est deux fois.’\textsuperscript{74}

The prescribed policies to achieve pluralism vary. While some stress the need for free access and choice in a competitive market, others suggest a more active role for the state in regulating the media to allow access to a broader range of information and ideas.\textsuperscript{75} The right policy (mix) varies depending on the type of media and their characteristics: ensuring competition may be easier for the written press due to the relative ease of starting a new publication than for broadcast media\textsuperscript{76}. This has led to a distinction between external and internal pluralism. External pluralism (pluralism of media), which characterises the written press, involves having a multiplicity of actors, which theoretically ensures diversity of viewpoints.\textsuperscript{77} Internal pluralism (pluralism of opinions or content pluralism), often imposed on broadcast media, involves having diverse information and opinions, which translates into

\textsuperscript{69} Maxwell McCombs and Sebastián Valenzuela, ‘The Agenda-Setting Theory’ [2007] Cuadernos de información 46.
\textsuperscript{72} Informationsverein Lentia and others v Austria [1993] ECtHR (Grand Chamber) App No. 13914/88; 15041/89; 15717/89; 15779/89; 17207/90.
\textsuperscript{74} Bernard Voyenne, Le droit à l’information (Aubier Montaigne 1970) 196.
\textsuperscript{75} van Hoboken (n 71) 347.
\textsuperscript{76} Voyenne (n 74) 204–205.
\textsuperscript{77} Berka and Tretter (n 68) 11.
a requirement to give enough coverage to different viewpoints. Media pluralism is especially important during elections. The next section will be devoted to RTI during elections, with focus on media regulation and pluralism during elections.

3.3. RTI during elections

3.3.1. Content and rationale of RTI in election periods

RTI is of great significance during election periods. Democratic elections require transparency, which is provided by RTI and requires freedom of expression. The 'right to vote cannot be exercised without the freedom to seek, receive and impart information if elections are to be genuine.' The right to political expression, which includes the freedom to seek, receive and impart information, is thus vital for elections to be ‘genuinely democratic’. Information allows voters to make an informed decision about the candidates and their platforms. RTI is important to ensure that both voters and candidates have information about the electoral process.

During elections, the right to seek, receive and impart information thus encompasses:

- Information about the electoral process and its integrity.
- Information about candidates, which implies that the state should remove obstacles to RTI that would hinder the ability of voters to make an informed choice.

This means that parties and candidates should be able to communicate their views, that the media should be able to report freely on electoral matters and that voters should receive all this information: ‘Ultimately, if the candidates cannot communicate or the media cannot report freely, then the electors will not have the information that they need. It is not just freedom of expression that will suffer, but the whole democratic process’.

This information can be imparted through different means:

---

78 ibid.
80 ibid 2.
81 ibid 6.
82 ibid.
83 Carver (n 73) 68.
84 UNCHR, ‘General Comment 25’ (n 45) para 12;25.
85 Carver (n 73) 68.
86 ibid 69.
• News coverage by broadcast or print media and under their direct editorial control. This involves a difficult balancing act between the need for impartiality and the protection of free editorial choice. Thus, the Venice Commission indicated that ‘the neutrality requirement applies to the electoral campaign and coverage by the media, especially the publicly owned media’87. The Council of Europe (CoE) likewise stressed that ‘that the fundamental principle of editorial independence of the media assumes a special importance in election periods’88 but also recommended ‘that the coverage of elections by the broadcast media should be fair, balanced and impartial’89. This reflects the special status of the broadcast media compared to print media (also recognised by the ECtHR90), as the Council recommends self-regulation or state intervention to ensure that ‘public service media and private broadcasters, during the election period, should be fair, balanced and impartial in their news and current affairs programmes, including discussion programmes such as interviews or debates’91. As always, when reporting about elections and political matters, ‘the safeguard afforded by Article 10 to journalists in relation to reporting on issues of general interest is subject to the proviso that they are acting in good faith in order to provide accurate and reliable information in accordance with the ethics of journalism’92.

• Special election programmes, such as debates between candidates, which can be assimilated to news programmes and subjected to the same requirements: ‘Where self-regulation does not provide for this, member states should adopt measures whereby public service media and private broadcasters, during the election period, should in particular be fair, balanced and impartial in their news and current affairs programmes, including discussion programmes such as interviews or debates’93.

• Directly by candidates and parties, through either free slots or paid advertising under the editorial control of the parties or candidates (political advertising). When free slots are

---

89 ibid.
90 Orlovskaya Iskra v Russia [2017] ECtHR (Third Section) App No. 42911/08 [109].
91 CoE (n 88) 4.
92 Orlovskaya Iskra v. Russia (n 90) [109].
93 CoE (n 88) 4.
provided, they should be on a non-discriminatory basis that ensures equal access to all candidates. However, various definitions of equal access co-exist (full equality vs proportional equality) and the ECtHR has held that no general right to be allocated airtime exists for political parties during election periods, provided that it does not lead to one party being denied any airtime while others are granted it.\textsuperscript{94} Regarding paid political advertising, practices vary widely between states and range from outright bans, to limit on expenditures, to more liberal regimes up to unlimited advertising.\textsuperscript{95} ECtHR case law has evolved over the years towards greater tolerance of restrictions on political advertising. While it had held in \textit{VgT v Switzerland}\textsuperscript{96}, \textit{TV Vest v Norway}\textsuperscript{97} and \textit{VgT v Switzerland (No. 2)}\textsuperscript{98} that bans on political advertising were violations of Article 10, it later recognised in \textit{Animal Defenders International v UK} that such bans may be justified to prevent undue influence by wealthy organisations and distortion of the democratic process\textsuperscript{99}. It specifically recognised that a blanket ban could be necessary to avoid ‘wealthy bodies creating many similar interest groups, thereby accumulating advertising time’\textsuperscript{100} and that it was not illogical for it to apply only to broadcast media (and not for instance the internet) due to the specificities of this medium.

- Voter education under the form of neutral material informing people of their rights and of electoral procedures.

3.3.2. Media regulation framework in France and Kenya: Application to elections

This section examines the media regulation framework that governs coverage of elections and communications related to political matters during elections. Both traditional media and social media will be studied, as regulation of social media, when it exists, often derives and expands upon pre-existing regulation of traditional media. Two countries were chosen: France, a European country known for its extensive media regulation apparatus, and Kenya, a developing country with hotly contested elections and a media regulation framework less protective of freedom of the press and RTI. These two countries recently held elections in which social media and controversial tactics were employed to (attempt to) influence the electoral process:

\textsuperscript{94} Partija ‘Jaunie Demokrāti’ and Partija ‘Mūsu Zeme’ v Latvia (Admissibility Decision) [2007] ECtHR (Third Section) App No. 10547/07, 34049/07 12–13.
\textsuperscript{96} VgT v Switzerland [2001] ECtHR (Second Section) App No. 24699/94.
\textsuperscript{97} TV Vest and Rogaland Fensjonistparti v Norway [2007] ECtHR (First Section) App No. 21132/05.
\textsuperscript{98} VgT v Switzerland (No2) [2009] ECtHR (Grand Chamber) App No. 32772/02.
\textsuperscript{99} Animal Defenders International v United Kingdom [2013] ECtHR (Grand Chamber) App No. 48876/08 [112].
\textsuperscript{100} ibid 122.
In France, hackers released two days before the second round of the 2017 presidential election a large collection of emails purported to have been writing by the Macron campaign team. These so-called ‘MacronLeaks’ then spread through social media.

The 2017 Kenyan Presidential election saw significant involvement by data firm Cambridge Analytica, also infamous for its role in the 2016 United States Presidential election during which it improperly obtained data on tens of millions of Facebook users to better target voters.

France

France is characterized by an extensive media regulation framework, which plays an important role in the period immediately preceding elections but also throughout the year. France has a two-tiered system that distinguishes between written press and broadcast (audio-visual) media. Currently, France does not have a press council to (self-)regulate the print media, although proposals to establish one have been put forward. The print media are thus subject only to the framework of the Loi du 29 juillet 1881 sur la liberté de la presse. By contrast, broadcast media are regulated by the Conseil Supérieur de l’Audiovisuel (CSA).

During elections, all types of media must abide by a set of fundamental rules. These include:

- Not publishing opinion polls on the eve and on the day of the election. This prohibition once extended to the week preceding the election, but was shortened to comply with the ruling of the Cour de Cassation on 4 September 2001 that it was incompatible with article 10 ECHR.

This prohibition applies to all media and individuals, including publications on social media.

---

107 Loi n° 77-808 du 19 juillet 1977 relative à la publication et à la diffusion de certains sondages d'opinion, Art. 11. https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000522846
108 Le Parisien [2001] Cour de Cassation (Chambre criminelle) 00–85.329 reported in Bull crim no. 170.
• Not publishing preliminary results before the last polling station has closed, through any means of communication\textsuperscript{109}
• Not publishing electoral messages by candidates on the eve and on the day of the election through any electronic means of communication\textsuperscript{110}.

Political ads are tightly regulated: there is a blanket prohibition on paid political advertising in broadcast media\textsuperscript{111}. Furthermore, all paid political advertising related to the election is prohibited regardless of media in the six months preceding the election, except for fundraising purposes.\textsuperscript{112} This restriction applies to internet banner ads\textsuperscript{113} and should also apply to promoted posts and ads on social media platforms, although no election has yet been invalidated due to illegal social media ads, despite anecdotal evidence of irregularities\textsuperscript{114}. Free slots are allocated to candidates to broadcast campaign communications, with more airtime allocated to parties represented in parliament\textsuperscript{115}. These campaign communications can be produced by the public broadcaster to ensure equal opportunities.\textsuperscript{116}

This reflects the intention to ensure pluralism. For print media, characterised by lower barriers to entry, this is done primarily through the protection of external pluralism\textsuperscript{117} by preventing excessive concentration in the daily newspaper sector and by regulating the distribution of newspaper to ensure access to a diversity of news sources (most notably through the \textit{Loi Bichet})\textsuperscript{118}. Thus, during elections, newspapers have no legal obligation to cover all candidates equally or to present issues in a balanced fashion, although individual journalists should still abide by professional ethics.

\textsuperscript{109} Code électoral (French Electoral Code) 1964 art. L. 52.2.
\textsuperscript{110} ibid art. L. 49-1.
\textsuperscript{111} Loi n° 86-1067 du 30 septembre 1986 relative à la liberté de communication (Loi Léotard) 1986 art. 14; Délibération no 2011-1 du 4 janvier 2011 relative au principe de pluralisme politique dans les services de radio et de télévision en période électorale 2011 pt IV-1.
\textsuperscript{112} Code électoral (n 109) art. L. 52-1.
By contrast, audio-visual media must guarantee “internal pluralism”. This is because there were historically fewer audio-visual media due to the scarcity of the airwaves\(^{119}\), rendering consumer choice insufficient to ensure access to a plurality of viewpoints. The CSA thus extensively regulates the operations of the broadcast media during elections, through its *Délibération du 4 janvier 2011 relative au principe de pluralisme politique dans les services de radio et de télévision en période électorale\(^{120}\)* and through additional recommendations published before each election. For instance, for the latest presidential elections, the guidelines\(^{121}\) distinguished three separate periods with corresponding obligations:

<table>
<thead>
<tr>
<th></th>
<th>Speaking time</th>
<th>Airtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 February 2017 – Day</td>
<td>Equity</td>
<td>Equity</td>
</tr>
<tr>
<td>before publication of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>official list of</td>
<td></td>
<td></td>
</tr>
<tr>
<td>candidates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publication of official</td>
<td>Equity</td>
<td>Equity, similar broadcasting</td>
</tr>
<tr>
<td>list of candidates – Day</td>
<td></td>
<td>conditions</td>
</tr>
<tr>
<td>before campaign</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beginning of official</td>
<td></td>
<td></td>
</tr>
<tr>
<td>campaign</td>
<td>Equality</td>
<td>Equality</td>
</tr>
</tbody>
</table>

The speaking time of both candidates and their supporters must be counted by the broadcast media to ensure equity (access in proportion to the importance of each candidate) or equality of treatment. Airtime devoted to each candidate is likewise counted. A report must be submitted periodically to the CSA to ensure that no unfairness can be identified.

The emergence of the internet led to a debate regarding the regulatory framework to which online services would be subject.\(^{122}\) It was decided in 2004 that online services did not share the fundamental characteristics of radio and television\(^{123}\). As such, most online platforms are excluded from the oversight of the CSA.

---

119 Barraud (n 117).
120 Délibération no 2011-1 (n 111).
123 See CSA, ‘La Définition Des Services Relevant de La Communication Audiovisuelle’ (CSA Conseil Supérieur L’audiovisuel) <http://www.csa.fr/Services-interactifs/Web-TV-Web-Radio/La-definition-des-services-relevant-de-la-
Kenya

General framework

Kenya has had a media framework that, perhaps paradoxically, has been described as both too lax and too draconian. The then-framework in 2008 could thus be described as ‘weak, irresolute and inadequate’\textsuperscript{124}, while simultaneously being ‘hostile to media and communication development’\textsuperscript{125} and tolerant of practices such as banning live streaming during the broadcast of election results.\textsuperscript{126}

Following the highly contested 2008 elections, the media regulation framework has been significantly overhauled since the adoption of a new Constitution in 2010. Unlike the previous Constitution (1963)\textsuperscript{127}, whose section 79 protected freedom of expression without explicit reference to freedom of the press and listed numerous restrictions, the 2010 Constitution\textsuperscript{128} incorporates a ‘Bill of Rights’ that distinguishes between freedom of expression\textsuperscript{129} and freedom of the media\textsuperscript{130}. It protects the ‘freedom and independence of electronic, print and all other type of media’, subject only to restrictions on hate speech, incitement to violence, propaganda for war and to the need to protect the rights and reputation of others. All such restrictions must be ‘reasonable and justifiable in an open and democratic society based on human dignity, equality and freedom, taking into account all relevant factors’\textsuperscript{131}. Likewise, the Constitution restricts state interference with the media, enshrines the independence and impartiality of state-owned media and entrusts media regulation to an independent body established by an Act of Parliament.\textsuperscript{132}

In 2013, two laws were enacted seemingly to give effect to these provisions by overhauling the media regulation framework: The Media Council Act\textsuperscript{133} and the Kenya Information and Communications (Amendment) Act\textsuperscript{134} (KIC Act). The first Act establishes the Media Council of Kenya as regulator of all media enterprises, journalists and other media practitioners.\textsuperscript{135} While this Act includes provisions

---

aimed at preventing abuses of journalistic freedom, such as by requiring journalists to ‘respect the personal dignity and privacy of others’ and to be ‘accurate and fair’, it also brought back much-abused restrictions such as national security and public morality considerations.\textsuperscript{136}

The Media Council Act is accompanied by a comprehensive Code of Conduct, which mandates that all media practitioners ‘present news fairly and impartially, placing primary value on significance and relevance’, ‘write a fair, accurate and an unbiased story on matters of public interest’, and ‘not publish a story that falls short of factual accuracy and fairness’. This Code of Conduct is included in the Second Schedule to the Act and is therefore not an act of self-regulation by the media but of legislative fiat. Failing to observe this Code of Conduct carries important financial consequences, including for individual journalists who can be fined following a complaint to the Complaints Commission.\textsuperscript{137}

The KIC Act deals primarily with broadcast media and establishes a similar body, the Communications and Multimedia Appeals Tribunal. Despite its name, this tribunal is not merely an appellate body, though it is also tasked with appeals related to decisions taken by the Complaints Commission. Rather, the Tribunal ‘is empowered to receive complaints from any person aggrieved by any publication or conduct of a journalist’\textsuperscript{138}, which creates significant overlap with the Complaints Commission. Inexplicably, although ‘the definition of ‘media’ in the KIC Act specifically excludes print and book publishing, […] the Tribunal’s authority extends to cover the print media’\textsuperscript{139}.

Internet media is regulated in the same fashion under the Media Council Act and the Information Act.\textsuperscript{140} A Computer Misuse and Cybercrimes Act 2018 was also recently passed and criminalises the online publication of false information.\textsuperscript{141}

Electoral periods

Regulation of the media during elections is not directly addressed by the Acts described above. However, subsidiary legislation giving further effect to these Acts has introduced regulatory rules. The bodies created under these Acts have also undertaken to offer guidance and binding rules in the period leading up to elections.

\begin{itemize}
\item \textsuperscript{136} ibid 4.
\item \textsuperscript{137} ibid 12.
\item \textsuperscript{139} ibid 345.
\item \textsuperscript{140} ibid 350.
\end{itemize}
The Publication of Electoral Opinion Polls Act\textsuperscript{142} has prohibited since 2012 the publication of opinion polls within five days of an election.

The Kenya Information and Communications (Broadcasting) Regulations (KICB)\textsuperscript{143} regulates broadcasting services (TV and radio). In addition to rules mandating balanced coverage of news at all times\textsuperscript{144} and especially when dealing with controversial issues\textsuperscript{145}, these Regulations require broadcasters during an election period to:

- provide equitable coverage and opportunities to registered political parties participating in an election and in particular to presidential candidates;
- ensure that the name of the political party or sponsor, if any, on whose behalf a broadcast is made, is announced, immediately before the commencement and immediately after such broadcast;
- ensure that the employees of a licensee who wish to be candidates for any elective position(s) resign from their employment with the licensee during polling period.

The programming code for free-to-air radio and television services\textsuperscript{146} further specifies that “Equitable opportunities in the access of unpaid airtime shall be given to candidates and political parties”\textsuperscript{147}, “No programme or sponsor shall be allowed to manifestly favour or oppose any candidate or political party”\textsuperscript{148} and that “The amount of airtime allotted to political propaganda and the rates to be charged for it shall be consistent to all parties and candidates”\textsuperscript{149}. Furthermore, all election propaganda must be identified as such and indicate the name of the candidate or political party that has paid for it\textsuperscript{150}.

The Media Council of Kenya likewise published guidelines for election coverage in preparation of the 2017 Presidential elections.\textsuperscript{151} These guidelines were agreed to by the most important stakeholders, including print media, and may be seen as elucidating the existing Code of Conduct. While not legally binding, the principles they contain could be used when a complaint is filed against a media institution.

\begin{itemize}
\item \textsuperscript{142} Publication of Electoral Opinion Polls Act, 2012 (Act No 39 of 2012).
\item \textsuperscript{143} The Kenya Information and Communications (Broadcasting) Regulations, 2009.
\item \textsuperscript{144} ibid 21.
\item \textsuperscript{145} ibid 24.
\item \textsuperscript{147} ibid 9.2.1.
\item \textsuperscript{148} ibid 9.2.2.
\item \textsuperscript{149} ibid 9.2.3.
\item \textsuperscript{150} ibid 9.2.5.
\end{itemize}
In practice, this regulatory framework and the Constitutional provisions have not always been sufficient to avert a return to problematic practices. In February 2018, several TV stations were thus forced off the air for several days, despite court orders, in order to prevent coverage of opposition rallies.\textsuperscript{152} Other sources of concerns have also been identified\textsuperscript{153}, including about the threat to freedom of expression and RTI that the new Cyber Crimes Act may represent\textsuperscript{154}.

### 3.4. Conclusion

This chapter has highlighted the importance of RTI both generally and during electoral periods. The study of two different media regulatory regimes has shown significant overlap in both goals and policy as regards the protection of pluralism and access to the media during elections. The importance of pluralism for democratic debate cannot be overstated and, in Europe at least, recent decisions appear to move towards (even) greater acceptance of restrictions aimed at ensuring a level playing field and at limiting the influence of moneyed interests. To protect pluralism and the right to receive diverse information from a variety of viewpoints, states often develop extensive media regulation frameworks, primarily for broadcast media. It is widely recognised that lifting barriers and adopting a \textit{laissez-faire} approach is not sufficient: positive intervention by the state is often necessary to protect RTI. As such, limiting certain aspects of RTI (e.g. the freedom to impart information) may be necessary to safeguard the broader objective of RTI (i.e. having an informed citizenry) and other aspects of it (e.g. the right to receive information, understood as the right to receive diverse information).

The emergence of online platforms has however been identified as a challenge to this existing regulatory framework, which largely predates the rise of Google, Facebook and similar actors.\textsuperscript{155} As the next chapter will discuss, such platforms now play an important role as gatekeepers of information, adopting an increasingly active role in filtering the news shown to their users. This raises important questions about the effect that these platforms have on RTI.

However, the overview of the media regulation frameworks of Kenya and France has also identified the need to be sensitive to disparities in the robustness of democratic institutions across countries. As the criticism surrounding the Cyber Crimes Act show, it cannot be assumed that states will always act

---


\textsuperscript{155} Inna Lyubareva and Fabrice Rochelandet, ‘Modèles économiques, usages et pluralisme de l’information en ligne’ [2017] Réseaux 9, 15–16.
in the best interests of their citizens. The traditional approach of imposing negative obligations upon
the state is thus still needed and care should be taken to avoid the chilling effect that can result from
well-intentioned attempts to regulate platforms that, in many countries, also serve as alternative
sources of information relied on by citizens when traditional media fail to inform the public. The last
chapter of this paper will return to this issue.
CHAPTER 4: ALGORITHMIC ACCOUNTABILITY

4.1. Introduction

Since the dawn of the computer age and the first forays into Artificial Intelligence (AI), concerns have been voiced regarding the dangers and ethical quandaries posed by AI. Alan Turing, widely considered the father of computing, thus remarked in 1951 that ‘it seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers. […] At some stage therefore, we should have to expect the machines to take control’\(^\text{156}\). These concerns often centred around the development of an ‘ultraintelligent’ machine whose intellectual capabilities surpassed those of human beings in all respects\(^\text{157}\), leading to doomsday predictions that rapidly percolated into popular culture, as evidenced by movies such as WarGames (1983) and The Terminator (1984) or, earlier, Stanley Kubrick’s 2001: A Space Odyssey (1968). The periodic re-emergence of such concerns tracks the ebb and flow of optimism about AI\(^\text{158}\), as the rapid progress witnessed during periods of ‘AI springs’ and ‘AI summers’\(^\text{159}\) never fails to reignite worries.

For the past decade\(^\text{160}\), we have been in such an AI summer, with the rise of machine learning and the expansion of AI into every area of human activity. The goals of these efforts are typically more modest: they do not aim to create ultraintelligent machines but to solve a specific problem, be it to recommend movies\(^\text{161}\), recognising cat faces\(^\text{162}\), or beating humans at Go\(^\text{163}\). However, they have given rise to new concerns as machine learning is increasingly used to develop systems that make life-altering decisions, such as classifying tumours\(^\text{164}\) or sentencing offenders\(^\text{165}\). This type of automated decision making has...

---


\(^{158}\) For an overview of AI history, see ibid 16–28.


\(^{160}\) Franklin (n 159).


\(^{164}\) David Capper and others, ‘DNA Methylation-Based Classification of Central Nervous System Tumours’ (2018) 555 Nature 469.

been criticised as ‘opaque’\textsuperscript{166}, incapable (or unwilling) to explain the factors that allowed the system to reach a given conclusion. This is undesirable because bias can unknowingly and unknowably be introduced\textsuperscript{167}, because it runs the risk of robbing individuals of their right to exercise their free will\textsuperscript{168} and because unexplainable decisions reduce social acceptability\textsuperscript{169}.

Consequently, significant work has been done toward identifying the problems with machine learning as it is currently practiced and implemented, the desired properties of a system based on machine learning, and the concrete steps that must be taken to bring about these favourable outcomes. This work has primarily revolved around the broad concept of ‘algorithmic accountability’ and involves several distinct types of actors:

1. Scholars in a variety of fields, most notably law, philosophy and media studies, who have sought to provide scholarly discussion of the issues raised by machine learning and algorithms and have been developed the key concepts in this debate.
2. (Quasi-)Official bodies (e.g. media regulators, judicial authorities), which often act as clearinghouses of previous research and have sought to summarize the issues and solutions identified by other scholars, usually with particular focus on their realm of competence. This chapter will primarily rely on reports published by French and British bodies:

   i) The 2014 report by the Conseil d’Etat on fundamental rights in the digital era, which addresses the issue of platforms and algorithms.\textsuperscript{170}
   ii) The 2017 report by the French privacy watchdog Commission Nationale Informatique et Libertés (CNIL).\textsuperscript{171}
   iii) The 2018 Villani Report published within the framework of the AI for Humanity initiative.\textsuperscript{172}

\textsuperscript{172} Villani (n 169).
iv) The 2017 Report by the research department of the CSA. ¹⁷³
v) A 2015 report by the Government Office for Science, providing a broad introduction to ethical issues surrounding machine learning.¹⁷⁴
vi) The UK privacy watchdog (ICO) 2017 Report on big data and machine learning.¹⁷⁵

3. ‘Popularisers’, such as Safiya Umoja Noble (‘Algorithms of Oppression’¹⁷⁶) and Cathy O’Neil (‘Weapons of Math Destruction’¹⁷⁷).

4. NGOs, such as the World Wide Web Foundation or the various entities that submitted comments in the context of the House of Commons Algorithms in decision-making inquiry¹⁷⁸.

This chapter will first provide an overview of the main challenges posed by algorithms and the platforms that use them. The concept of algorithmic accountability, which aims to address these concerns, will then be explored and analysed. Finally, the proposals put forward to make algorithms accountable will be discussed, with special focus on the implications for the subject matter of this paper, visibility and invisibility of online news.

4.2. The problem with algorithms and online platforms

Although current debate has primarily centred on the role of online platforms during elections and on ‘fake news’, the growing importance of algorithms and of the platforms using them was already under scrutiny before Cambridge Analytica became a household name. The rapid increase in algorithmic decision-making led many observers to identify significant sources of concern. Some of these concerns applied to all uses of algorithms, while others were more specific to filtering and ordering platforms such as Facebook, YouTube and the Google search engine.

4.2.1. Loss of human control

Fears of loss of control over automated decision-making (ADM) systems have been voiced since the early days of the computer age. This is reflected for instance in data privacy laws, which restrict fully

¹⁷³ CSA Lab (n 168).
automated decision making.\textsuperscript{179} The perspective of automated systems making decisions indeed raises questions about the loss of accountability and human agency that the widespread use of such systems would cause.\textsuperscript{180} This fear is most prominent in French reports, as evidenced by the title of the CNIL Report (‘comment permettre à l’homme de garder la main?’) and early legislative attempts to address this problem\textsuperscript{181}. Other reports have identified similar issues, noting for instance that automating tasks can increase psychological distance and therefore cause human beings to take less consideration of the human impact of the decisions taken.\textsuperscript{182}

### 4.2.2. Bias and discrimination

Computers and algorithms are often falsely perceived as objective decision-makers under the mistaken belief that ‘the computer is always right’.\textsuperscript{183} However, it is now recognised that algorithms can embody prejudice and lead to discrimination, for instance because the data used to train the algorithm reflected (human) biases.\textsuperscript{184} Thus, ‘Hidden biases in both the collection and analysis stages present considerable risks’\textsuperscript{185}. For instance, a ProPublica report identified that software widely used to predict the risk of reoffending appeared to be biased against African-American offenders, overstating their likelihood to reoffend.\textsuperscript{186} Likewise, Google ads were especially likely to suggest that an individual might have an arrest record when names commonly given to African-Americans were entered.\textsuperscript{187} Many other examples have been identified, for instance by Safiya Umoja Noble\textsuperscript{188} and Cathy O’Neil\textsuperscript{189}.


\textsuperscript{180} Commission nationale de l’informatique et des libertés (CNIL) (n 171) 26–31; Conseil d’Etat (n 170) 234–236.

\textsuperscript{181} See Article 10 Loi n° 78-17 du 6 janvier 1978 relative à l’informatique, aux fichiers et aux libertés.


\textsuperscript{185} Crawford (n 183) para 2.


\textsuperscript{187} Latanya Sweeney, ‘Discrimination in Online Ad Delivery’ (2013) 56 Communications of the ACM 44.

\textsuperscript{188} Noble (n 176).

\textsuperscript{189} O’Neil (n 177).
4.2.3. Opacity

Concerns about bias and discrimination is further compounded by the opaque nature of many algorithms. First, users are typically provided with little information about the inner workings of algorithms and about how they were applied in a given situation. For instance, a user cannot know why Google has ranked search results in a particular way or how exactly Facebook prioritised the content of their newsfeed. Not only were many users unaware until recently that their newsfeed was curated, such decisions are also made invisible by design as users never see the many posts that the algorithm decided not to display. In other domains, algorithms may act more visibly (e.g. when applying for a travel authorisation such as ESTA) but still provide no justification for the decision made. The ubiquitous nature of algorithms has led Frank Pasquale to speak of a ‘black box society’ in which

Credit raters, search engines, major banks, and the TSA take in data about us and convert it into scores, rankings, risk calculations, and watch lists with vitally important consequences. But the proprietary algorithms by which they do so are immune from scrutiny, except on the rare occasions when a whistleblower litigates or leaks.

Second, certain advanced machine learning techniques (primarily deep learning) are intrinsically opaque and therefore generate results that cannot be explained even by those who employ them. Thus, ‘the lack of interpretability in such algorithmic decision-making systems therefore threatens the ability of decision-makers to account for their systems’. Often no-one is able to explain why and how an algorithm reached its decision. This may not be an issue when there is gold standard against which results can be evaluated, but may not be acceptable when it leads to opaque decisions about more subjective matters.

---

190 Villani (n 169) 46.
192 Tufekci (n 9) 209.
195 Binns (n 184) 11.
4.2.4. Liability issues

As autonomous systems (e.g. self-driving cars) become more widespread, issues relating to legal liability also become more pressing. This is not a wholly new problem, as similar issues arose with the rise of the personal computer. Nissenbaum for instance already identified that responsibility was difficult to assign when a computerised system malfunctioned.\(^\text{197}\) Even when legal safeguards exist, they are not always effective, for instance because of the nature or the number of actors involved.\(^\text{198}\) This makes it difficult to establish the level of human involvement (if any).\(^\text{199}\)

Furthermore, who is responsible for holding algorithms accountable is often unclear:

> To whom do we appeal? What bodies govern artificial intelligence, and where does the public raise issues or lodge complaints with national and international courts? These questions have yet to be fully answered.\(^\text{200}\)

4.2.5. Status of platforms

The legal framework regarding internet services predates to a large extent the rise of algorithmic platforms such as Facebook. These platforms thus do not fit tightly within the categories set out in the e-commerce Directive (2000/31/CE) and domestic law\(^\text{201}\):

- They are not intermediary service providers (‘mere conduits’, caching services or hosting providers), as that they do not confine themselves to a passive role but actively filter and order information. This is in line with some ECJ jurisprudence on the matter\(^\text{202}\), although the status of platforms such as Facebook remains somewhat controversial as cases before national courts show\(^\text{203}\).
- They are unlikely to be considered editors or publishers under domestic law, as they do not typically exercise the same type of editorial control.

---


\(^\text{198}\) Commission nationale de l’informatique et des libertés (CNIL) (n 171) 46; Nissenbaum (n 197) 29.


\(^\text{200}\) Noble (n 176) 28.

\(^\text{201}\) Conseil d’Etat (n 170) 219–222.

\(^\text{202}\) L’Oréal and Others [2011] ECJ C-324/09 [107–117].

\(^\text{203}\) CG v Facebook Ireland Ltd & McCloskey [2016] Northern Ireland Court of Appeal 54.
Thus, while it is advantageous for platforms to benefit from the same protections as more passive hosting companies (notably limited liability and not obligation to proactively monitor all content\(^{204}\)), the current framework is ill-suited to deal with the new challenges posed by algorithmic platforms.

### 4.2.6. Filter bubbles, microtargeting, lack of diversity and pluralism

Personalisation of content is at the core of many online platforms. Google filters and ranks results to provide users with relevant information. Facebook populates users’ newsfeed with the content that it deems most relevant to that particular user, for instance news articles\(^{205}\). YouTube, Netflix and Spotify can recommend media based on previous viewing and listening habits.

Users can be shown widely different content based on their characteristics and proclivities, with no way to see content that the platform has hidden. Individuals might be confined to ‘bubbles’, creating a segregated internet and reducing exposure to unfamiliar ideas\(^{206}\). Although evidence is inconclusive regarding the present reality of such bubbles\(^{207}\), the rising influence of platforms warrants caution. While users may indeed not live in a bubble as they consume other media offline, there may still be little pluralism in the news that they receive through platforms that are increasingly important.

Microtargeting individuals to present a tailored message to them, for instance for the purpose of electoral propaganda could lead to increasingly fragmented political discourse, reducing common grounds between citizens\(^{208}\).

A related concern is that the most common filtering and recommendation algorithms favor content that is similar to previously consumed content. For instance, YouTube will recommend content that matches what the user has previously watched. Thus, while users have theoretically access to a wide range of content, they are quickly pigeonholed into a much more restricted media diet\(^{209}\). Little attention is also paid to cultural diversity\(^{210}\).


\(^{206}\) Conseil d’Etat (n 170) 234–236; Commission nationale de l’informatique et des libertés (CNIL) (n 171) 34–38.

\(^{207}\) Frederik Zuiderveen Borgesius and others, ‘Should We Worry About Filter Bubbles?’ (2016) 5 Journal on Internet Regulation 16.

\(^{208}\) Commission nationale de l’informatique et des libertés (CNIL) (n 171) 36–37.

\(^{209}\) CSA Lab (n 168) 13.

\(^{210}\) ibid 15.
4.2.7. Spread of disinformation

Several elections have recently been characterized by disinformation, often referred to by the moniker of ‘fake news’. For instance, in the 2016 United States Presidential Election, ‘fake news’ was shared on Facebook at least 38 million times, although the impact of such news remains under dispute.\(^{211}\)

Beyond election periods, there are concerns that ‘by creating repeated distortions impacting citizens’ perceptions of events, [disinformation] can give rise to deep-seated misinformed beliefs and cause ‘significant harm’\(^{212}\), as elections are contested based on issues that are put on the agenda before campaigns start.

Particularly concerning is the indication that online disinformation tends to spread faster and reach more people than reliable news, and this is especially true of political disinformation.\(^{213}\)

4.2.8. Censorship

Online platforms have gained significant power and acquired near total control over what users see and do not see. This implies that they have the ability to determine what ideas and individuals will be allowed to spread the views on their platforms. To protect themselves against legal liability in countries with hate speech laws or in the case of copyright infringement, many platforms have deployed automated takedown systems that can identify and remove content that they deem illegal or in contravention of their terms of service.

This has led to criticism that such systems can stifle freedom of speech by aggressively identifying content as offensive or illegal. For instance, Facebook famously deleted the ‘napalm girl’ photo because of the nudity that it contains.\(^{214}\) Similar complaints have been leveled against YouTube’s automated copyright infringement detection system.\(^{215}\)

---


4.2.9. Application to the visibility and invisibility of online news on Facebook

Most of the issues outlined in this section apply to Facebook and the visibility and invisibility of news on this platform.

**Loss of human control:** Algorithms automatically curate the newsfeed, with little user control over parameters, no human editor, no human quality check (except indirectly through opaque measures of website quality). This can lead to feelings of dispossession when users realise that their newsfeed is wholly different from their acquaintances even though they never explicitly requested such personalisation.

**Bias, discrimination and filter bubbles:** This issue is central to debates about the role of Facebook in society and in electoral contests in particular. It is possible to run sponsored posts on Facebook using criteria that target specific demographics and such issue has only recently began to be tackled seriously. As noted above, Facebook presents each user with an individualised newsfeed that leads to filter bubbles. During elections, users will tend to see news from sources aligned with their preferred candidate or ideological proclivities, and see neither ads from the other candidates nor hear opposing viewpoints (or only infrequently).

**Opacity:** Limited information is available regarding the ranking criteria. When information is made available, it is either incomplete (e.g. why a particular sponsored post was shown) or too generic to be useful (generic explanations regarding the philosophy of the ranking algorithm, with no details about the weight given to each factor).

**Liability issues:** In theory, there is a corporate structure that could be held accountable and liable in case of algorithmic harms. However, recent experience shows little effective accountability. Although algorithmic harms were committed, no redress seems forthcoming and the blame is likely to rest primarily or solely on third-parties said to be in breach of the platform’s terms of service.

**Status of platforms:** Although Facebook plays an important role during elections and is a major source of information, it benefits to a large extent of legal provisions that were meant for more passive intermediary service providers, such as hosting providers. However, their role is much more active: in addition to curating the newsfeed, Facebook also embeds employees inside political campaigns to allow them to better use its targeted ad functionalities.\(^\text{216}\) The concerns regarding ‘fake news’ has also

led to Facebook assuming an increasingly active editorial role, taking a more explicit stance about which content is valuable (‘meaningful social interactions’\(^{217}\)) and which news sources are reliable and useful (‘high-quality’ sources, local content, etc.)\(^{218}\). Much of these decisions are made in an non-transparent fashion, based on debatable value judgements (e.g. that longer articles are better\(^{219}\)). While this may reduce the spread of disinformation, an important problem, it raises questions about censorship, as how accurate and useful a news source is judged could have an important impact on its viability given the large influence of Facebook.

### 4.3. Algorithmic accountability: Definition and principles

The concept of algorithmic accountability has been developed to address these issues. Algorithmic accountability is a concept whose exact contours remain to be fully defined. Accountability generally involves the giving of reasons following an action taken against an individual who has a legitimate claim to demand an account and is usually followed by sanctions.\(^{220}\) Accountability thus has two related aspects: providing reasons for one’s actions and taking responsibility for them. Applied to algorithms, this has translated into a requirement for transparency, defined as ‘the disclosure of information about algorithms to enable monitoring, checking, criticism or intervention by interested parties’\(^{221}\), and into a need to hold automated decision makers responsible for their actions\(^{222}\). Thus, algorithmic accountability ‘includes an obligation to report, explain, or justify algorithmic decision-making as well as mitigate any negative social impacts or potential harms’\(^{223}\).

Algorithmic accountability is motivated by the need to retain control over the outcomes of algorithmic processes. As Nicholas Diakopoulos noted, ‘We’re living in a world now where algorithms adjudicate more and more consequential decisions in our lives. […] Algorithms, driven by vast troves of data, are the new power brokers in society.’\(^{224}\) As a consequence, ceasing control is necessary. Control can

---


\(^{222}\) Binns (n 184) 2.


\(^{224}\) Diakopoulos, ‘Algorithmic Accountability Reporting’ (n 184) 2.
however take several forms. First, it is possible to prevent problems *ex ante* to avoid the realisation of harms. Second, redress can also come *ex post*, by identifying those responsible and making them liable.

Several proposals have been put forward to translate algorithmic accountability into more concrete principles. FAT/ML identified responsibility, explainability (including to non-technical users), accuracy (being aware of sources of errors and uncertainty), auditability (being able to probe to workings of an algorithm) and fairness (lack of discriminatory or unjust impacts) as key characteristics of an accountable algorithm.\(^\text{225}\) Likewise, the Association for Computing Machinery (ACM) proposed seven principles for algorithmic transparency and accountability: awareness (being aware of potential biases and of their impact), access and redress (mechanisms allowing affected individuals to seek redress), accountability (institutions should be held responsible, even when results cannot be explained in detail), explanation (providing explanations regarding how decisions were made), data provenance (describing how the training data was collected), auditability (models and decisions should be stored to be audited), and validation and testing (models should be thoroughly tested and the results made public).\(^\text{226}\)

IEEE also published a broader draft report on ethically-aligned design in AI, which identified general principles: respect for human rights (including cultural diversity), prioritizing human well-being, accountability (which includes liability and record-keeping to apportion blame and avoid potential harm), transparency, awareness of the potential for misuse.\(^\text{227}\)

This overview shows substantial agreement and overlap about the fundamental principles underpinning algorithmic accountability, although differences do exist. For instance, accountability seems to be sometimes restricted to liability, in which case its other component parts are nevertheless included in a broader ethical framework (e.g. by the ACM and IEEE). Conversely, accountability has also sometimes been conflated with and equated to transparency, which has increasingly been criticised as insufficient to achieve accountability on its own.\(^\text{228}\)

In this section, a broad definition of algorithmic accountability, rooted in the need to retain control over algorithms both *ex ante* and *ex post*, will be used. This follows Binns’s remark that algorithmic accountability is often an ‘umbrella term for a variety of measures, including transparency, auditing

\(^{225}\) FAT/ML (n 223).


\(^{228}\) World Wide Web Foundation (n 184) 10.
and sanctions of algorithmic decision-makers. This definition will therefore include all principles that have been identified as necessary or valuable in ensuring that algorithms are kept in check, act in the public interest, and are operated according to legal and ethical standards. Care will also be taken to differentiate between principles to be safeguarded (e.g. explainability, fairness) and the means (technical or not) to achieve these desired outcomes (e.g. auditing algorithms). Many discussions and proposals, such as those listed above, fail to make this crucial distinction. The remainder of this section will focus on what algorithmic accountability aims to achieve, while the next section will discuss how they should be achieved.

4.3.1. Transparency

Transparency has been central to efforts to achieve algorithmic accountability, as it was seen by early exponents of this concept as the ideal means to ensure effective control over algorithms. This stemmed from the centrality of transparency in public discourse regarding effective ways to make institutions accountable to the public by allowing all to examine the motives behind peoples’ actions by accessing more information about the decision-making process. Transparency is therefore seen as a potent tool to promote social accountability and trust by reducing ‘information asymmetry’ between governments and their citizens, between businesses and their customers and between news platforms and their audiences.

This transparency ideal is rooted in the assumption that ‘being able to see how a system works equates knowing how to govern it’ and that observing and knowing a system leads to effective control. This provided the impetus for the many Freedom of Information (FOI) legislations and for the inclusion of this transparency ideal in various ethical codes of conducts and policies in medicine, journalism and law.

When applying this model to algorithms, a central question has been the degree to which transparency is required and the effectiveness of transparency. The goal of transparency is to allow users to see how an algorithm works, presumably in order to assert control over them and/or effect change. This includes

---

229 Binns (n 184) 2.
230 Diakopoulos and Koliska (n 221) 811.
232 ibid 13.
234 Mike Ananny and Kate Crawford, ‘Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability’ (2016) 20 New Media & Society 973, 973.
235 ibid 975.
236 ibid 976.
for instance knowing what data was used to make a decision and how the algorithm used this data to make a decision, in a format that can be understood by non-technical users. Diakopoulos argues for extensive transparency requirements, including the criteria used to rank, the parameters used, descriptions of the data used to train the model, the accuracy of classifications, and the thresholds used to identify similar items.

Despite such proposals, there is currently no consensus regarding which justifications are adequate, especially since standards may differ across individuals and social groups and between users and owners of platforms. There is also a need to balance transparency and other imperatives, such as the necessity to protect trade secrets and prevent misuse of the system, or to protect national security.

While most sources and reports identify transparency as an important principle it has become increasingly apparent that transparency is not sufficient to guarantee algorithmic accountability. Indeed, being able to see how an algorithm works is often not enough, as it is also necessary to understand the inner mechanics of the algorithm. Furthermore, there is no guarantee that knowledge about the workings of an algorithm will translate into greater control over it. More recent scholarship, including those that still advocate for algorithmic transparency, have thus also recognised that there are limitations to the pure transparency approach to algorithmic accountability. Annany and Crawford outline ten limitations of the transparency-centred approach to algorithmic accountability:

i. Transparency can be disconnected from power: Revealing problems does not necessarily lead to accountability, as there may be no structures to hold those responsible to account.

ii. Transparency can be harmful: ‘radical transparency’ is not always desirable as privacy is also important and vulnerable minorities need some degree of anonymity to be protected.

---

237 Conseil d’Etat (n 170) 237–239.
238 Commission nationale de l’informatique et des libertés (CNIL) (n 171) 51.
239 Diakopoulos, ‘Algorithmic Accountability’ (n 233) 411.
240 Binns (n 184) 2.
241 ibid 6.
242 Pasquale, ‘Restoring Transparency to Automated Authority’ (n 191) 237.
243 Diakopoulos, ‘Algorithmic Accountability’ (n 233) 403.
244 Karanasiou and Pinotsis (n 199) 183.
246 Annany and Crawford (n 234) 978–982.
iii. Transparency can occlude: making too much information available makes it impossible to identify truly useful information.\(^{248}\)

iv. Transparency can create false binaries: calls for transparency can result in a binary view that opposes complete secrecy and complete openness, which can both be problematic.

v. Transparency promotes neoliberal models of agency: transparency often works in an individualistic fashion, placing the burden of seeking information and leveraging it on individuals.

vi. Transparency does not necessarily build trust.

vii. Transparency entails professional boundary work: apparent openness, e.g. making data available, can be compatible and encourage more obscure types of gatekeeping.

viii. Transparency privileges seeing over understanding: Seeing may not be enough to understand how an algorithm works. Experimenting with it may be necessary.

ix. Transparency has technical limitations: Computer systems are often not too complicated to be understood by individuals lacking technical skills, but can also be incomprehensible even to their creators due to their design. This is especially true of deep learning systems which make decisions whose logic cannot be explained.

x. Transparency has temporal limitations: for instance, adaptive systems have interfaces, settings and capabilities that change over time, and what exactly should be made transparent is not always clear: ‘Online machine learning systems update their decision rules after every query, meaning that any disclosure will be obsolete as soon as it is made.’\(^{249}\)

Ananny and Crawford rightly highlight that focusing only on the need to make the algorithm transparent can occlude the fact that an algorithmic system relies on both code and on practices and norms that are enforced by a variety of actors.\(^{250}\) Thus, to hold the algorithmic system accountably requires that all components be held accountable, rather than merely shining light on a single part of the system. It is for these reasons that scholars have argued that transparency should constitute the first step towards accountability and that more concrete actions should be taken to ensure algorithmic accountability. This involves looking at more substantial matters and at how algorithms should (be

---


\(^{251}\) Pasquale, ‘Restoring Transparency to Automated Authority’ (n 191) 254.
allowed to) perform rather than merely revealing how they do perform in the hope that it will automatically guarantee good behaviour.

4.3.2. Loyalty of algorithmic systems

The French Conseil d’Etat promotes the concept of loyalty of online platforms which requires algorithms acting in the best interests of their users. For instance, platforms should not secretly improve the ranking of some content because of commercial relationships between the platform and the originators of this content.

This concept of loyalty can be interpreted in an individualistic fashion as a requirement for algorithms and platforms to faithfully provide users with what they consider best for them. In that perspective, Facebook should show users the posts that is most relevant to them. However, it can also be interpreted in a more collective fashion, recognizing that what is best for individuals is not necessarily socially optimal. As Cathy O’Neil notes, ‘what’s efficient for political campaigns is inefficient for democracy’. What is efficient for users may also be inefficient for democracy or society as a whole. For instance, while a user may like a newsfeed that does not challenge their pre-existing beliefs, perhaps such a filter bubble would be detrimental to democratic debate.

There is therefore a certain tension between requirements of individual and collective loyalty. This might be resolved by means of an appeal to ‘public reason’, ‘universal rules [that] must be justifiable on grounds that are suitably public and shared by all reasonable people in the society, and without appeal to beliefs that are controversial’. In a society where the need for media diversity is widely accepted, the necessity of limiting individual loyalty in order to recommend more diverse media content may for example be shared by all reasonable people.

This principle of loyalty can be extended to cover various types of improper behaviour that run counter to algorithmic accountability, such as discrimination or censorship. Tackling discrimination is indeed important as this is included in most ethical guidelines proposed for algorithmic systems and in most reports about algorithmic accountability. From an individual perspective, discrimination runs counter to loyalty as most users do not use the platform in order to be discriminated against. From a collective perspective, even if some users (e.g. announcers) might prefer the algorithm to discriminate,

---

252 Conseil d’Etat (n 170) 273.
251 Commission nationale de l’informatique et des libertés (CNIL) (n 171) 49.
255 Binns (n 184) 7.
256 Conseil d’Etat (n 170) 237–239; Villani (n 169) 116; Information Commissioner’s Office (n 175) 20; World Wide Web Foundation (n 184) 8–9.
this would also be contrary to the public reason principle as non-discrimination is a core values of our societies, enshrined in domestic legislation and international instruments. There is therefore a duty to ensure that algorithms are free from bias.

4.3.3. Due process and procedural regularity
Discrimination involves unfairly treating individuals differently based on personal characteristics such as gender, race, sexual orientation, etc. An important requirement for algorithms is therefore that they ensure procedural regularity so that the algorithm is applied consistently and not in a selective fashion (due process). Procedural regularity means that ‘each participant [or user] will know that the same procedure was applied to her and that the procedure was not designed in a way that disadvantages her specifically’.

4.3.4. Compliance with legal and ethical requirements
As algorithms do not operate in a legal vacuum, it is important that they comply with relevant legislation (e.g. anti-discrimination laws). Algorithms should also obey the ethical principles that are foundational in the society in which they are used (in addition to ethical requirements specific to ensure algorithmic accountability, such as those described here). This includes respect for human rights.

4.3.5. Responsibility and liability
Being able to assign blame in the case of algorithmic harms is essential since it ensures that there will be answerability for failures or harms caused by their use, which creates a strong incentive to minimise harms and provides the first basis for assigning sanctions and compensation for victims where necessary (algorithmic justice). This means that the defence that ‘the algorithm did it’ must be rejected and that there must be clear rules to make companies liable.

Responsibility also stresses the role of human beings. Algorithms can reflect the biases of their creators and of society and their creators must thus take responsibility for their creation. Algorithmic accountability requires those who create algorithms bear the risk posed by biased algorithms and to reassert control over their algorithms.

257 Kroll and others (n 249) 22.
258 Karanasiou and Pinotsis (n 199) 184.
259 The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (n 227).
260 Nissenbaum (n 197) 27.
261 Or ‘It’s the Computer’s Fault’, see ibid 34.
Responsibility also implies that users have access to those who designed and operate algorithms in order to air their grievances and suggest changes.

4.3.6. Overview of principles

The five principles described subsume the principles found in the three reports detailed above. Table 1 shows the principles used in the present paper and their correspondence in previous reports. ‘(Means)’ indicates that this principle is best seen as a means towards another end: in such cases, the principle was linked to the broader principle presented in this paper. Auditability was thus linked to both transparency (as it involves making available the means to understand the workings of the algorithm in an *ex ante* fashion) and to responsibility (as it allows blame to be assigned after algorithmic harm has been committed, *ex post*).

<table>
<thead>
<tr>
<th>FAT/ML</th>
<th>Transparency</th>
<th>Loyalty</th>
<th>Due process</th>
<th>Compliance</th>
<th>Responsibility and liability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responsibility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Explainability</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy <em>(means)</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Auditability <em>(means)</em></td>
<td>X (ex ante)</td>
<td></td>
<td></td>
<td>X (ex post)</td>
<td></td>
</tr>
<tr>
<td>Fairness</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ACM</th>
<th>Transparency</th>
<th>Loyalty</th>
<th>Due process</th>
<th>Compliance</th>
<th>Responsibility and liability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Awareness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Access and redress</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Accountability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Explainability</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data provenance</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auditability <em>(means)</em></td>
<td>X (ex ante)</td>
<td></td>
<td></td>
<td>X (ex post)</td>
<td></td>
</tr>
<tr>
<td>Validation and testing <em>(publication)</em></td>
<td>X</td>
<td></td>
<td></td>
<td>X (awareness of the problem)</td>
<td></td>
</tr>
<tr>
<td>Human rights</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human well-being</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accountability and liability</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transparency</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Awareness of misuse</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 1. Overview of proposed algorithmic accountability principles*

### 4.3.7. Application to the visibility and invisibility of online news on Facebook

At the close of this section, these five principles can now be applied to assess how the treatment of content on Facebook fares compared to the ideal of algorithmic accountability.

Transparency remains sorely lacking. For organic content, no information is given about why a particular post was shown and in that rank. Neither the data used to make that decision, nor the parameters of the algorithm are displayed. How the various parameters that Facebook has identified in high-level descriptions of its algorithm are combined is not revealed. While some of the design choices (e.g. prioritising meaningful interactions) have been made public, how they translate and are weighed against other considerations is unclear. Nor do users have meaningful control over these design choices. While ads provide more information, there is still insufficient transparency. For instance, while it has been shown that many political ads in the 2016 US Presidential election originated from Russia\(^{262}\), this information was not disclosed. Comparing the targeting criteria found in released ads with what is disclosed to users shows that much of the criteria remain hidden.\(^{263}\)

Given this lack of transparency, there is no guarantee that newsfeeds are fairly populated or that due process and procedural regularity is observed. This also raises questions regarding the loyalty of Facebook. It is unknown whether the content shown receives this ranking because it is ‘best’ for users or because it serves some further purpose for Facebook (increasing retention or time spent on Facebook, optimising revenue, making users more pliable to ads, etc.). This is especially true when adopting a more collective understanding of loyalty, as there is also no way to determine whether the

---


algorithm has a positive impact on democracy and on the public sphere. The filter bubble effect casts doubt on this.

**Compliance** with ethical principles is increasing as a result of public outcry regarding elections. However, it is still in its infancy and much of the work still seems to happen behind closed doors and using self-defined and self-imposed conceptions of good corporate stewardship. While this has raised new questions regarding censorship and user control over the platform, there has nevertheless been some progress, for instance through GDPR compliance. Yet, it remains to be seen whether this will translate into greater responsibility and liability.

4.4. **Methods to achieve algorithmic accountability**

Many proposals have been put forward to tackle these problems outlined above and achieve algorithmic accountability.

4.4.1. **Fostering transparency**

Transparency is a fundamental requirement of algorithmic accountability and there is widespread agreement that greater transparency is necessary. Online platforms should display at least information about the general workings of their algorithms, in a format understandable by the general public.\(^{264}\)

This does not imply that all information should necessarily be made available as there may be legitimate reasons to retain some opacity (e.g. to prevent abuse, protect national security but also to protect trade secrets and encourage innovation\(^{265}\)). It is therefore important not to set up a false dichotomy between total secrecy and total transparency. Nevertheless, algorithms should still carry an explainability requirement\(^ {266}\) and users should obtain explanations without having to make an explicit, burdensome request.\(^ {267}\) If secrecy is important, a trusted third-party could examine the algorithm.\(^ {268}\)

There has thus been significant discussion around the amount of information that should be made transparent and what form such disclosure should take. Providing the source code of the system may be of limited use but is widely considered insufficient since most users are not technical enough to understand it\(^ {269}\) and static source code is uninformative when dealing with adaptive systems. It has thus been suggested that at the very least the goal, purpose and intent of the algorithm should be

\(^{264}\) Diakopoulos, ‘Algorithmic Accountability’ (n 233) 411; Commission nationale de l’informatique et des libertés (CNIL) (n 171) 51.

\(^{265}\) Government Office for Science (n 174) 15–16; CSA Lab (n 168) 18.

\(^{266}\) Villani (n 169) 124.

\(^{267}\) Commission nationale de l’informatique et des libertés (CNIL) (n 171) 56.


\(^{269}\) Diakopoulos, ‘Algorithmic Accountability’ (n 233) 411.
Likewise, ‘Many experts and commentators have suggested that transparency is necessary to ensure accountability: being clear which algorithms are used, which parameters, which data, to what end, will be necessary to determine whether the technology has been used responsibly’\textsuperscript{271}. Being transparent about the data used to train the system has also been identified as critical, as has the features of the model being used (variables chosen, weighing of features, etc.).\textsuperscript{272} It is likely that the degree of transparency will depend on what the algorithm is used for.

If there is some consensus regarding the need for greater transparency, there remains debate about how this goal should be achieved. While some have explicitly called for a transparency mandate, at least in some contexts\textsuperscript{273}, as found for instance in the GDPR, others have been more cautious, considering that regulators and public authorities should primarily act as facilitators rather than resort to heavy-handed intervention in a rapidly changing landscape\textsuperscript{274}.

Thus, scholars have suggested technical means that could help achieve transparency by showing how the algorithm was used in a user-friendly manner\textsuperscript{275}. An interface should be provided to users to allow them to obtain more information about the algorithm\textsuperscript{276} and to interact with it. Transparency is not static but can better be achieved through interactive tweaking that allows for better understanding of the algorithm.\textsuperscript{277} For example, users could obtain an explanation regarding why specific content was shown to them but also alternative recommendations by interactively changing the parameters fed to the algorithm, their weighing, etc.

Algorithmic literacy has also been identified as necessary to allow the general public to better understand how algorithms work.\textsuperscript{278}

\begin{flushright}
\textsuperscript{270} Diakopoulos and Koliska (n 221) 8.
\textsuperscript{272} See Diakopoulos and Koliska (n 221) 9 for a comprehensive list of elements that can be made transparent.
\textsuperscript{273} Conseil d’Etat (n 170) 24.
\textsuperscript{274} CSA Lab (n 168) 18.
\textsuperscript{275} Diakopoulos, ‘Algorithmic Accountability’ (n 233) 411.
\textsuperscript{276} Diakopoulos and Koliska (n 221) 12,13,15.
\textsuperscript{277} Commission nationale de l’informatique et des libertés (CNIL) (n 171) 43.
\textsuperscript{278} Noble (n 176) 26; Osonde A Osoba and William Welser IV, An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence (Rand Corporation 2017).
\end{flushright}
4.4.2. Auditing algorithms

Because transparency is considered insufficient or is not always provided by online platforms, it has been suggested that algorithms should be auditable. As Sandvig et al. note, ‘Rather than regulating for transparency or misbehavior, we find this situation argues for ‘regulation toward auditability.’”

This can be done through reverse engineering (interacting with the algorithm and changing the input provided to infer its workings from the visible output), as efforts show. However, it is desirable to make such an endeavour easier by having APIs made available by the platforms themselves. This would allow either public authorities or civil society groups to investigate algorithms.

Auditing would allow concrete means of enforcing algorithmic accountability: algorithms could be rated based on the results of auditing to allow users to know whether an algorithm acts in their best interests, auditing could identify bias and allow civil society to campaign for change or public authorities to mandate changes.

Unlike some more static forms of transparency, auditing takes into account the fact that algorithms should be evaluated in the context in which they are used.

4.4.3. Accountability by design

It has been argued that the need for accountability should be baked in and taken into account when designing algorithmic systems. This can be dubbed ‘accountability by design’ and, within that framework, ’ex ante efforts could include monitoring or prescribing the algorithm’s design features and principles for example, carefully selecting training data or initial weights so that they are consistent with legal or ethical constraints.

---

279 Conseil d’Etat (n 170) 301.
281 Commission nationale de l’informatique et des libertés (CNIL) (n 171) 58.
282 Diakopoulos, ‘Algorithmic Accountability’ (n 233).
283 Villani (n 169) 117.
284 ibid 118.
285 CSA Lab (n 168) 18.
286 Sandvig and others (n 280).
287 Joshua Kroll, 2016|Algorithmic Accountability and Featured4 Comments, ‘Accountable Algorithms (A Provocation)’ (Media Policy Project, 10 February 2016) <http://blogs.lse.ac.uk/mediapolicyproject/2016/02/10/accountable-algorithms-a-provocation/> accessed 14 May 2018. Kroll uses this phrase in a more restricted sense. It will be used here as an umbrella term for all approaches that aim to ensure that algorithms are designed with accountability in mind.
288 Karanasiou and Pinotsis (n 199) 184.
To ensure that ethical standards are respected data scientists should be given training in ethics and that professional standards and self-regulatory bodies be created for this emerging field.\textsuperscript{289} Ethics board could also be set up to examine algorithms as they are developed\textsuperscript{290} and codes of conduct could be written.\textsuperscript{291}

AIAs have been proposed to evaluate the risks and consequences of use of an automated system before such a system is made available.\textsuperscript{292} Like impact assessments in other fields (e.g. environmental protection), such AIAs would evaluate algorithms to assess the risks that they pose and identify mitigating measures, making the risk/benefit balance clearer. FAT/ML has proposed a detailed template for a social impact statement for algorithms, to be used at the design stage, the pre-launch stage and the post-launch stage to evaluate algorithms. This template includes systemic examination of the five principles for algorithmic accountability identified by FAT/ML.\textsuperscript{293}

To remedy filter bubbles, introducing some degree of randomness to preserve serendipity has been suggested\textsuperscript{294}, as has showing users content widely different from what the algorithm predicts the user will like.\textsuperscript{295} Involving users more and giving them more say over criteria used can also foster accountability by design.

Finally, research efforts are seen as necessary to develop techniques that are more consonant with the need for accountability, such as the Explainable AI research program\textsuperscript{296}.

\subsection{4.4.4. Creating a new status for online platforms}

In light of the discrepancy between the current legal framework and the specific characteristics, the creation of a new category of ‘online platforms’ has been suggested, with the requirement that such platforms act loyalty towards their users.\textsuperscript{297}

\begin{footnotesize}
\begin{enumerate}
\item\textsuperscript{289} Government Office for Science (n 174) 15; Villani (n 169) 119–120; Commission nationale de l’informatique et des libertés (CNIL) (n 171).
\item\textsuperscript{290} Information Commissioner’s Office (n 175) 88.
\item\textsuperscript{293} FAT/ML (n 223).
\item\textsuperscript{294} Urbano Reviglio, ‘Serendipity by Design? How to Turn from Diversity Exposure to Diversity Experience to Face Filter Bubbles in Social Media’ in Ioannis Kompatsiaris and others (eds), \textit{Internet Science - 4th International Conference, INSCI 2017, Thessaloniki, Greece, November 22-24, 2017, Proceedings} (Springer 2017) <https://doi.org/10.1007/978-3-319-70284-1_22>.
\item\textsuperscript{295} CSA Lab (n 168) 13.
\item\textsuperscript{296} D. Gunning (n 166).
\item\textsuperscript{297} Conseil d’Etat (n 170).
\end{enumerate}
\end{footnotesize}
4.4.5. Creating stronger legal accountability mechanisms

Clearer lines of responsibility and accountability should be created within companies. Users should know who designed the algorithm and an executive should be made responsible for algorithmic decisions and tasked with accounting for the workings of the algorithm (Chief Algorithm Officer)^{298}. This approach is consonant with the requirement to have human beings be the final decision makers for momentous decisions.

4.4.6. Introducing oversight mechanisms and overhauling the regulatory system

There is currently no real oversight mechanism for algorithms beyond sporadic efforts by sector-specific regulators to tackle the issues that algorithms raise in their domain of responsibility. There have therefore been calls to establish oversight bodies to better regulate algorithms. Two models have been proposed:

- An ‘FDA for algorithms’^{299}, which would work primarily *ex ante*. This independent public agency would assess, track and measure the impact of a proposed algorithmic system over time. Before an algorithmic system could be deployed, it would have to be thoroughly evaluated.\(^{300}\) This would ‘prevent the introduction of certain algorithms in the market until their safety and efficacy have been proven through evidence-based pre-market trials’^\(^{301}\).  

- An ‘NTSB for algorithms’^{302}, which would work primarily *ex post*. After algorithmic harms have been committed and identified, this agency could investigate algorithmic accidents and assign liability accordingly, making recommendations or imposing measures to prevent further harm.

In addition to these suggestions to set up a dedicated body tasked with algorithmic regulation, there have also been calls to further empower existing regulatory bodies to regulate algorithms that affect their sector or to expand the mission of privacy watchdogs.^\(^{303}\)

4.4.7. Application to the visibility and invisibility of online news on Facebook

There is little doubt that greater *transparency* is required. Static transparency (e.g. releasing the source code) would be ineffective and too intrusive in light of the objective and the need to prevent individuals from gaming the system. Facebook should at the very least make a comprehensive, generic explanation

\(^{298}\) Government Office for Science (n 174) 15.  
\(^{300}\) Dillon Reisman and others, ‘ALGORITHMIC IMPACT ASSESSMENTS’: 22.  
\(^{301}\) Tutt (n 299) 91.  
\(^{303}\) Conseil d’Etat (n 170) 301.
of its system available, describing the value judgements made, how they relate to each other and other contents, and the most important factors used in deciding which news content to show and whether to show news content or other types of content.

For ads, especially political ads, all targeting criteria should be easily accessible in order for users to know why they were targeted. This would also raise awareness about how certain characteristics can be used to infer political opinion, race, etc.

In terms of interface, implementing a way for users to manipulate the algorithm and ‘burst’ the filter bubble would be ideal. This should apply to both organic content and ads. While there is already a plan to make a searchable archive of political ads, these should be shown alongside the complete, precise targeting criteria to allow for public scrutiny.

Because transparency will necessarily be limited, auditing facilities need to be made available while protecting users’ privacy. These should at least be made available to national authorities.

To promote accountability by design, an ethics board could be set up, involving all stakeholders (users, journalists, academics, etc.), which would review the principles underlying the algorithm and proposed changes to it, in order to guarantee that the platform remains loyal to its users. The deliberations of this board would be made public, which would also increase transparency. Likewise, an impact assessment should be conducted on the ranking algorithm, again with public disclosure of its results.

Filter bubbles could be avoided through more serendipity and a commitment to preserving diversity of viewpoints by modifying the algorithm to present a less biased reflection of the political landscape (a further discussion of the proposal in light of existing media regulation during elections is reserved for the next chapter).

The evolving nature of Facebook should be recognised: from a platform used primarily to share personal content, it has evolved into an important electioneering and news platform, which has asserted increasing control over what should be shown. This calls for the creation of a new category of actor that would be better suited to the nature of such platforms. This will also be discussed in the next chapter, along with the legal accountability mechanisms associated with this new category and the need for an overhaul of the regulatory framework.

CHAPTER 5: RIGHT TO INFORMATION, ALGORITHMIC ACCOUNTABILITY, VISIBILITY

5.1. Right to information and the (in)visibility of online content

As Chapter 3 has stressed, the right to information, like other rights, must not be ‘theoretical and illusory’ but ‘practical and effective’. It is increasingly recognised that limiting state interference may not be enough to make the right to information effective, as inaction can also prevent individuals from fully enjoying their right to information. Failure to foster the environment necessary for the enjoyment of this right, failure to intervene in relationships between private parties can also interfere with the effective realisation of the right to information.

Chapter 4 has shown the extent to which online platforms select which content to display, in a non-transparent manner, with a tendency to display content similar to what users are predicted to want based on their history and personal characteristics. This has led to concerns that such platforms create ‘filter bubbles’ that artificially restrict the breadth of information shown to users.

As such, these practices raise questions regarding their compatibility with the right to information. It is however necessary to distinguish between worries about the potential effects of such filter bubbles on society and (legal) claims that online platforms infringe on the right to information. This distinction is important to separate aspirational claims (what platforms should do in an ideal world, what users should require them to do) and legal claims (which existing obligations online platforms breach, requiring state intervention). When examining claims of third-party infringement on the right to information, the courts have indeed had to pay detailed attention to the circumstances of the case and the broader context in which the restriction occurred. An important factor was whether alternative means of obtaining the same information were available. News, the type of content most often cited when examining filter bubbles, are typically available through other channels, which would seem to limit the ability of online platforms algorithms could prevent individuals from accessing news. However, other considerations point to an important role of platforms in determining what information users receive:

305 Khurshid Mustafa (n 61) [45].
• Online services such as Google and Facebook account for 70% of traffic received by publishers.\textsuperscript{306}

• A greater share of the population now uses online platforms than broadcast media to access\textsuperscript{307}. Younger generations seem especially likely to rely on social news for information.

• The impact of online platforms also occurs \textit{ex ante}, as publishers pursue news stories that will be more valuable on social media (‘create buzz’) and better fit the criteria used by online platforms.\textsuperscript{308}

• Important stories, especially concerning marginalised groups or in countries with restricted press freedom, increasingly originate and spread on online platforms before being picked up by traditional media.

Thus, the significance of such platforms cannot be reduced to statistics about media consumption. Preventing access to the only available source of information, as in \textit{Khurshid}, is certainly a violation of RTI, but it does not automatically follow that no role for the state exists in safeguarding RTI when different sources of information are available or when certain sources have only a limited audience.

First, pluralism has been recognised as an important characteristic of a democratic society, which creates place for positive actions by the state to create an environment favourable to pluralism. This means that while platforms may not directly infringe on each individual user’s right to information, their actions may still run counter to the principles that underpin RTI.\textsuperscript{309} Second, a more systemic approach must often be taken. If online news are quantitatively less important than commonly assumed, newspapers are also an increasingly rare source of information.\textsuperscript{310} Yet, this does not imply that measures aimed at fostering external pluralism in the written press could be discontinued without adverse effects or that restrictions on the activities of print journalists are more tolerable now that few people rely on newspapers. The different types of media do not work in isolation: a story first published in print media (or on a social media platform) and initially seen by few readers may be amplified by broadcast media (or a social media platform) far beyond its original audience.


\textsuperscript{309} Eskens, Helberger and Moeller (n 38) 26.

\textsuperscript{310} Newman and others (n 307) 10.
Relying on the availability of alternative news sources to minimise the importance of online platforms would also fail to recognise that these platforms are different in kind from other media. As, in *Khurshid*, newspaper and radio programmes were not equivalent to television broadcasts, content available on online platforms is not necessarily equivalent to that accessible through other sources. An important characteristic of online platforms is their reliance on user-generated content. While users sometimes only share pre-existing content, they often add their own comments and interact with other users, which helps shape their opinion and contributes to public debate.

Thus, concerns about filtering and ranking (and attempts to regulate platforms) also implicate the right to impart information. The non-transparent algorithms used to populate newsfeeds are of relevance not only to consumers of information but also to producers, as they determine whether the information that they sought to impart will actually reach users. While platforms have given a voice to many users who would not have had access to traditional media, they also tend to reduce pluralism by giving more importance to a limited number of news outlets\(^{311}\). Recent announcements that online platforms will attempt to distinguish between trustworthy and untrustworthy news outlets\(^{312}\) raise questions regarding whether this would not further run counter to the objective to foster RTI. Such efforts also raise the spectre of viewpoint discrimination, as they may make it more difficult for less mainstream opinions to be shared, thus restricting rather than expanding pluralism.

Recognising the *sui generis* nature of online platforms is key to effectively regulating them. The combination of user-generated content and algorithmic editorial choices makes them different from previous types of media. This is best illustrated using the case study of content visibility on Facebook during elections. Many have legitimately voiced concerns that the filter bubble phenomenon had a detrimental impact on democracy and that the unbalanced, personalised newsfeeds of Facebook prevent users from hearing the diverse viewpoints necessary to form an informed opinion about candidates and issues. However, examining the instruments used to alleviate this problem in traditional media shows that reasoning by analogy will not be sufficient:

- For print media, the traditional policy is to foster external pluralism to ensure access to a wide variety of viewpoints. However, applying this perspective to social media could lead to the conclusion that no problem exists. Indeed, as noted above, the information is often available elsewhere. Furthermore, there is much potential pluralism on Facebook. Each user can directly

\(^{311}\) Dario Compagno and others, ‘La reconfiguration du pluralisme de l’information opérée par les réseaux socionumériques’ (2017) 205 Réseaux 91.

receive updates about electoral and political matters from Breitbart or Buzzfeed, Fox News or CNBC, Le Figaro or Mediapart, Daily Nation or The Standard, which all maintain a presence on the platform. The problem is thus that users do not seek out information that is available but that they do not wish to see or are simply unaware exists. The more individualistic framework used to ensure access to information contained in print media is thus ill-suited to tackle the new challenges posed by online platforms. It does, however, correctly tend to focus on intermediaries such as newspaper distributors and newspaper agents, for instance by imposing that a wide range of newspapers be made available to customers.

- The internal pluralism enforced in broadcast media is closer in philosophy to what is typically envisioned. However, the analogy also breaks down at some point, as there is an infinity of ‘channels’ that could be created on Facebook (each user can create pages) and the scarcity rationale often used to justify internal pluralism does not directly apply. Nor is it clear how a system like that used to ensure equality or equity of coverage between candidates could be extended to social media platforms. This is not only because of practical difficulties (is it feasible to monitor all posts to ensure balance? What should be done when a user chooses to have online like-minded friends and news sources?) but also because such attempts would potentially be greater breaches of RTI than the problem that they intend to solve. Unlike broadcast media, which features professional content by journalists and candidates or their representatives, social media content is shared by users. Attempts to ensure more balance in newsfeeds cannot thus be as coercive as the regime applied to broadcast media, as this would likely constitute a disproportionate interference with the right to impart information enjoyed by users.

Algorithmic accountability may provide such a less stringent regime that can ensure non-discrimination, more pluralism and more transparency while preserving users’ freedom of expression. Rather than attempting to determine what an ideal Facebook newsfeed should look like, this framework can focus on procedural aspects and on principles that should be respected when determining which content to show. The next section will sketch out what such a framework could be, for both general issues relating to online platforms and more specifically for the visibility of online content during elections.
5.2. Fostering algorithmic accountability to ensuring a practical and effective right to information on online platforms

This section will provide concrete proposals for the regulation of online platforms, rooted in the principles of algorithmic accountability found in Chapter 4 and mindful of the need to preserve and foster RTI described in Chapter 3 and analysed in the previous section.

The keystone of this framework is the recognition that the specific nature of online platforms requires them to be afforded a special legal status, with rights and obligations that better correspond to their characteristics. The first subsection will thus provide an overview of how online platforms should be regulated in general. These general principles and rules will ensure that platforms always serve the interests of their users (including during election periods) and, by clarifying their status, will enable further regulation during elections. It is indeed easier to impose obligations on a well-defined and objectively delineated category of actors, which reduces the risk of unfairly targeting specific platforms while under-regulating others. The second subsection will turn specifically to content visibility during elections, outlining measures that could tackle this issue. Finally, the last subsection will adopt a more international perspective, highlight the necessity to take into account local sensibilities and challenges when discussing the need for regulation. This will also serve as a general conclusion.

5.2.1. Making platforms accountable

The first step in making platforms more accountable is to recognise that the current framework does not reflect the evolution of the internet landscape. Social media platforms are different in kind from the previous types of actors and cannot be assimilated to either mere technical providers (such as hosting companies) or publishers (such as website owners). It is therefore necessary to create a new category of actors, as already put forward by the French Conseil d'Etat in 2014.

This new category of ‘online platforms’ would be defined as an online service that serves content generated or shared by individual users and presents them according to editorial choices (including in terms of ranking) under at least partial control of the service, regardless of whether these editorial choices are the result of human or algorithmic decisions. This definition (to be further refined in order to cover all relevant services) would be included in an overhauled version of the e-commerce Directive, which would create some homogeneity among member states while giving room for local transposition to better reflect the needs of each country. This European definition would serve as an anchor to give national regulators a clear category on which to exert their powers, such as by imposing measures to foster cultural diversity in countries that pursue this objective.
At this level, only broader obligations would be imposed, leaving again a margin of appreciation regarding the precise means through which these obligations should be realised. It would here be important to include generous exemptions for actors whose size do not warrant the imposition of extensive regulation. The need for algorithmic accountability is most pressing for platforms that have acquired a systemic nature and have an important impact on the general public. Imposing onerous regulations on fledging start-ups with a limited customer base risks stifling innovation, which would run counter to the objective of reining in platforms such as Facebook as it would limit the emergence of potential competitors. Thresholds in terms of size (e.g. SME vs large enterprise) should thus be included when defining obligations and compliance mechanisms.

The two main obligations should be those of loyalty and transparency. Loyalty would include the need to treat all users (whether in their role as consumers or creators of content) equitably, in a non-discriminatory fashion, while always acting in their best interests (including collective interests such as the protection of pluralism). This is similar in spirit to the client protection measures contained in article 24 of MIFID II and would likewise be further elucidated by national authorities and regulators. For the largest actors, there should be an obligation to involve all stakeholders before major changes to the algorithm, such as a decision to deprioritise certain content. This could take the form of an ethics board that would verify that the principle of loyalty is not compromised. As the largest platforms often have a near-monopoly position in their market, users must be given an opportunity to comment on such proposed changes, with the output from such consultations taken into account by the ethics board. These ethics boards could also decide on preventive measures such as requiring data scientists to receive training in ethics. At periodic intervals, an outside body (e.g. an FDA for algorithms) could verify that the platform indeed operates consistently with the principle of loyalty.

Transparency would require platforms to publish comprehensive information about the general workings of their platform, including a detailed list of the various criteria and weightings used to display and rank content. This information would also be made available, unless technical reasons render it impossible and with authorisation from regulatory authorities, on the platform for each displayed item. To protect trade secrets and prevent gaming of the algorithm, not all information would be available. However, there should be a way for users to request to obtain more information if they can demonstrate a compelling reason (e.g. because they suffered an algorithmic harm). The framework put in place for access to public documents, could serve as an inspiration and national bodies could be created to give their opinion on whether the online platform should comply with a request. Such bodies could also serve as intermediaries and provide users with an answer while preserve the necessary confidentiality.
The liability and accountability of online platforms would also be reviewed. To make them more accountable, a human executive would have to be designated as the point of contact and the person accountable in case of suspected breach of these principles or of other obligations. However, the platforms would retain the limited liability that they currently enjoy, which also protects users against aggressive filtering. To ensure genuine accountability, national regulatory authorities would be empowered to regulate platforms on both a transversal and a thematic basis. This means that regulators tasked with specific domains would be able to regulate platforms active in these domains but that there would also be the possibility to create a regulator of all online platforms (to avoid regulatory overlap, it may be advantageous to give a choice between either of these two options).

Finally, competition authorities should scrutinise the market power exercised by these online platforms. One issue has always been precisely delineating the relevant market to assess dominant positions. Clarifying the status of online platforms may alleviate this problem by providing a clear category of actors that need to be taken into account when measuring market dominance.

5.2.2. Fostering the right to information during elections

The regulation of platforms to foster RTI during elections should first be based on a strong distinction between sponsored content and user-generated content (a further distinction may be made between true user-generated content and professional content such as news articles shared by users – this distinction will not be made here).

**Sponsored content**

Sponsored content should be treated according to the same rules that apply to broadcast media. The concerns linked to the broadcasting of political ads on TV and radio apply to social media platforms, perhaps even more acutely. Indeed, online platforms further blur the distinction between advertising and normal content and allow announcers to micro-target users. Countries that ban political advertising in audiovisual media should thus extend this ban to online platforms. When free airtime is given to political candidates, the same obligation should be imposed on social media platform, along with strict rules aimed at ensuring a level playing field (non-discriminatory targeting, size limitations, etc.).

Where political ads are allowed, rules governing the charging of non-discriminatory fees should apply. This may pose problems for platforms, as they typically charge for ads using an auction process that therefore gives different value to different target audiences. This has led to claims that political candidates can currently be charged widely different fees based on who they need to target.\(^{313}\) While

---

this may be similar to phenomena in broadcast TV (where the fees might vary depending on the program during which the ad is shown, which may likewise correlate with demographics), regulators would need to pay close attention to this and set clear rules.

Regardless of whether sponsored content is paid or freely provided according to law, it would be advantageous to remove sponsored political content from the newsfeed to clearly distinguish it from normal content. A link would be displayed alongside each ad allowing users to know why they were targeted, to know who paid for the sponsored post, and to see all other political ads that have been run during the campaign.

These proposals should not be problematic in light of RTI, as it has now been established in Europe that states can regulate or ban political advertising to limit the influence of moneyed interests and that online platforms can convincingly be shown to pose the same (and even greater) risks as broadcast media.

**User-generated content**

Regulating the display of user-generated content poses greater challenges in terms of RTI. On the one hand, it appears that news and online content personalisation ‘may enable or hinder the exercise of this largely institutionally protected right’\(^\text{314}\), with many stressing the hindering of this right. On the other hand, mandating that user-generated content be shown or not, or imposing extensive rules, might well run afoul of the right to impart information. Regulation should not consist of potentially dangerous attempts to social engineer what an ideal Facebook newsfeed should look like or empower platforms to become the arbiters of truthfulness and genuine pluralism. It should rather empower users and restrain the power of platforms, in keeping with the general principles found in the previous subsection.

This would require an even greater degree of transparency when showing political content, with a detailed explanation of the reasons why a particular post was shown (or hidden).

A key principle should be to supplement rather than replace: pluralism should not be enforced through illusory and unworkable quotas inspired by broadcast TV regulation but through user-friendly ways to explore all the content available on the platform. This would first take the form of customisable filters allowing users to interact with the algorithm and see the influence that their demographic characteristics and actions have had on the contents of their newsfeeds. Such a requirement could be included in the general rules governing the operation of online platforms, as it is relevant beyond political content during elections. However, such a facility would still be limited, on Facebook at least,

to content that friends have shared on the platform. This may therefore not do much to reduce filter bubbles, as users are likely to have many friends of the same political persuasion.

To go further, alternative political content from other professional sources (candidate, pressure group, news media) with publicly accessible posts should be easily accessible whenever a user has chosen to share this type of content. This would allow users to quickly see alternative perspectives, as Google News already allows, while preserving the ability of users to show their friends their chosen content. This would also allow users to quickly gauge the reliability of a piece of information by assessing whether many news sources report it.

This proposal fosters pluralism and increases the amount of information received by users while limiting the encroachment on the freedom of the platform and of the originating user, as content can still be shown according to the ranking determined by the platform. Alternative content does not replace but supplements the information currently shown, which minimises interference. Such measures may be an act of self-regulation by online platforms or be contained in soft law originating from regulators as they work out the implications of loyalty and transparency for political contents during elections and beyond.

However, stricter rules are necessary regarding current attempts to filter and rank the newsfeed using measures of journalistic quality. A good case can be made that such attempts should be banned as incompatible with the principle of pluralism inherent in the obligation to be loyal to users. Large private actors should not become the arbiters of journalistic integrity. There are clear risks that such attempts would unduly privilege large media outlets, which runs counter to the push to adopt a functional definition of journalism that would cover new actors such as bloggers and citizen journalists. Quality should be enforced in a less intrusive manner, according to the principle that information should supplement rather than replace. For instance, general information about the ideological characteristics of a news source could be shown, along with an ability to explore the same issue from other ideological perspectives.

This proposal should not be problematic as it is widely accepted that the state plays an important role in fostering pluralism, which this measure would seek to do. It would also be proportionate to its intended aim, as giving powerful actors the right to censor media outlets interferes with the right to receive and impart information to such an extent that banning this practice is the only appropriate measure. Finally, it merely supplements rather than replaces existing information.

The regulatory authorities, whether electoral authorities or the new online platform regulator, should issue periodic guidelines regarding the conduct of platforms in the upcoming election. This would
allow rules to be refined to match the latest advances in technology and the actual behaviour of the platforms. One requirement would be to conduct an algorithmic impact assessment prior to elections to identify potential challenges and remedy them before the beginning of the campaign. This would include an assessment of the risk of biased newsfeeds that unduly favour one candidate or one set of issues relative to the content that could potentially be shown to different categories of users. The philosophy here would be one of procedural regularity, ensuring that the same treatment is reserved to content from different perspectives.

5.3. Protecting freedom of speech and the right to information worldwide

The previous two subsections were written in an international perspective, giving measures that can *mutatis mutandis* apply worldwide (e.g. by replacing the overhaul of the e-commerce Directive with relevant legislative changes). However, it is important to pay due attention to local specificities. This requires taking into account two principles:

- The need for awareness and integration of measures into the local regulatory framework.
- The need to foster respect for universal principles, among which RTI, for instance through curbing censorship.

This implies that regulating online platforms could not in practice be done similarly in France as in Kenya. It must be remembered that the two impetuses for regulation in the public discourse have often been the fear of foreign influence in elections and the phenomenon of ‘fake news’. While many had been working on algorithmic accountability, it is only after revelations surrounding these two threats that the need for greater regulation was taken seriously.

This is not without danger in environments that are more inimical to freedom of the press. As noted in Chapter 3, the current Kenyan landscape is already characterised by repeated attempts to curb freedom of the press. Thus, the need to protect journalists and citizens against state interference may there carry more importance than issues surrounding the working of online platforms. Such platforms have indeed been used in environments with more limited freedom of the press to raise awareness about issues that were not or could not be discussed in traditional media. It is important to keep this in mind when proposing more extensive regulation of online platforms, as the tools and legal framework developed to protect users against the excessive private power of platforms may well be misused by states with less pure motives. Stressing the risks of foreign influence and fake news too much may legitimate action against civil society organisations and international NGOs under the pretext of foreign influence, or against (citizen) journalists under the guise of combatting fake news.
This explains why the present proposals have steered clear of more heavy-handed regulation, promoting instead a general framework that impose primarily obligations aimed at empowering users. This, rather than utopian attempts to achieve perfect balance in newsfeeds, may well be what is needed to ensure that the right of information becomes a reality worldwide.

Enshrining in law the category of online platforms is especially important to draw attention to the special nature of such services and to protect the rights of their users. While the previous sections primarily focused on how this change would help protect users against platforms, it may also protect users against state restrictions on RTI by clarifying under the ambit of which regulatory authorities (if any) these platforms should fall. Regulating by analogy, which ambiguity about the status of platforms encourages, risks imposing inappropriate obligations on platforms and their users. As this paper draws to an end, the Kenya Film Classification Board (KFCB) has for instance just announced that it would treat vloggers who upload video content on online platforms as filmmakers, requiring them to hold a licence and to pay a myriad of fees.315 Proper categorisation of platforms and of their users is a necessary first step towards limiting such regulatory overreach.

The closing words of this paper are words of caution. For all their ills, social media platforms are also potent catalysts of freedom of expression, allowing millions to express their opinions in hostile media environments. Their power must be exercised for the benefits of their users, avoiding both the tyranny of platforms and the more familiar tyranny of illiberal states. Care should be taken not to stifle RTI through heavy-handed regulation that ignores local contexts. The problems posed by platforms are worldwide and demand a response that is mindful to the needs and specific challenges of all their users, regardless of frontiers.

BIBLIOGRAPHY

Conventions, Treaties and Statutes


Code électoral (French Electoral Code) 1964


Délibération n° 2011-1 du 4 janvier 2011 relative au principe de pluralisme politique dans les services de radio et de télévision en période électorale 2011

International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171 (ICCPR)


Loi n° 78-17 du 6 janvier 1978 relative à l’informatique, aux fichiers et aux libertés

Loi n° 86-1067 du 30 septembre 1986 relative à la liberté de communication (Loi Léotard) 1986

Publication of Electoral Opinion Polls Act, 2012 (Act No 39 of 2012)

Recommandation n° 2016-2 du 7 septembre 2016 du Conseil supérieur de l’audiovisuel aux services de radio et de télévision en vue de l’élection du Président de la République 2016


The Constitution of Kenya, 2010 (CoK)

The Kenya Information and Communications (Amendment) Act, 2013 (Act No 41A of 2013)

The Kenya Information and Communications (Broadcasting) Regulations, 2009

The Media Council Act, 2013 (Act No 46 of 2013)

Universal Declaration of Human Rights (adopted 10 December 1948) UNGA Res 217 A (III) (UDHR)
Cases

Ahmet Yildirim v Turkey [2012] ECtHR (Second Section) App No. 3111/10
Animal Defenders International v United Kingdom [2013] ECtHR (Grand Chamber) App No. 48876/08
Board of Education v Pico [1982] Supreme Court of the United States 457 U.S. 853
Cengiz v Turkey [2015] ECtHR (Second Section) App No. 48226/10 and 14027/11
CG v Facebook Ireland Ltd & McCloskey [2016] Northern Ireland Court of Appeal 54
Dink v Turkey [2010] ECtHR (Second Section) App No. 2668/07
Gaweda v Poland [2002] ECtHR (Former Section I) App No. 26229/95
Handyside v UK [1976] ECtHR App No. 5493/72
Informationsverein Lentia and others v Austria [1993] ECtHR (Grand Chamber) App No. 13914/88; 15041/89; 15717/89; 15779/89; 17207/90
Kenedi v Hungary [2009] ECtHR (Second Section) App No. 31475/05
Khurshid Mustafa and Tarzibachi v Sweden [2008] ECtHR (Third Section) App No. 23883/06
Le Parisien [2001] Cour de Cassation (Chambre criminelle) 00–85.329 reported in Bull crim no. 170
L’Oréal and Others [2011] ECJ C-324/09
Magyar Kétfarkú Kutya Párt v Hungary [2018] European Court of Human Rights 201/17
Orlovskaya Iskra v Russia [2017] ECtHR (Third Section) App No. 42911/08
Partija ‘Jaunie Demokrāti’ and Partija ‘Mūsu Zeme’ v Latvia (Admissibility Decision) [2007] ECtHR (Third Section) App No. 10547/07,34049/07
Sdružení Jihočeské Matky v Czech Republic [2006] ECtHR (Fifth Section) App No. 19101/03
Stanley v Georgia [1969] SCOTUS 394 U.S. 557
The Sunday Times v United Kingdom [1979] ECtHR App No. 6538/74
Thomas v Collins [1945] SCOTUS 323 U.S. 516
TV Vest and Rogaland Pensjonistparti v Norway [2007] ECtHR (First Section) App No. 21132/05
VgT v Switzerland [2001] ECtHR (Second Section) App No. 24699/94
VgT v Switzerland (No2) [2009] ECtHR (Grand Chamber) App No. 32772/02
X and Church of Scientology v Sweden [1979] ECtHR App No. 7805/77
Youth Initiative for Human Rights v Serbia [2013] ECtHR (Second Section) App No. 48135/06
Articles, Books and Documents


Ananny M and Crawford K, ‘Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability’ (2016) 20 New Media & Society 973


Association for Computing Machinery US Public Policy Council (USACM), ‘Statement on Algorithmic Transparency and Accountability’ <https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf> accessed 13 May 2018


Bullinger M, ‘Freedom of Expression and Information: An Essential Element of Democracy’ 28 German Yearbook of International Law 88
accessed 1 May 2018

Capper D and others, ‘DNA Methylation-Based Classification of Central Nervous System Tumours’ (2018) 555 Nature 469

accessed 18 March 2018


CoE, ‘Recommendation CM/Rec(2007)15 of the Committee of Ministers to Member States on Measures Concerning Media Coverage of Election Campaigns (Adopted by the Committee of Ministers on 7 November 2007 at the 1010th Meeting of the Ministers’ Deputies)’  


Compagno D and others, ‘La reconfiguration du pluralisme de l’information opérée par les réseaux socionumériques’ (2017) 205 Réseaux 91

Conseil d’Etat, Le Numérique et Les Droits Fondamentaux (La Documentation française 2014)  
<http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/144000541.pdf> accessed 6 May 2018


CSA, ‘Durée et mode de repartition du volume horaire consacre a la diffusion des emissions lors des campagnes audiovisuelles officielles’  
<http://www.csa.fr/content/download/217357/583889/version/1/file/Dur%C3%A9e%20et%20mode%20de%20r%C3%A9partition%20du%20volume%20des%20missions%20des%20campagnes%20audiovisuelles%20officielles.pdf> accessed 18 May 2018

CSA Lab, ‘Le Rôle Des Données et Des Algorithmes Dans l’accès Aux Contenus’  
<https://www.cnil.fr/sites/default/files/atoms/files/csaroledesdonneesetdesalgorithmesdanslaccesauxcontenus.pdf> accessed 5 May 2018

D. Gunning, ‘Explainable Artificial Intelligence (Xai) : Program Update’  

——, ‘Algorithmic Accountability’ (2015) 3 Digital Journalism 398

——, ‘Enabling Accountability of Algorithmic Media: Transparency as a Constructive and Critical Lens’, *Transparent Data Mining for Big and Small Data* (Springer, Cham 2017) <https://link.springer.com/chapter/10.1007/978-3-319-54024-5_2> accessed 5 May 2018


Dwyer T and Martin F, ‘Sharing News Online. Social Media News Analytics and Their Implications for Media Pluralism Policies’ (2017) 5 Digital Journalism 1080


Frederik Zuiderveen Borgesius and others, ‘Should We Worry About Filter Bubbles?’ (2016) 5 Journal on Internet Regulation 16


Kramer ADI, Guillory JE and Hancock JT, ‘Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks’ (2014) 111 Proceedings of the National Academy of Sciences 8788


——, ‘Modèles économiques, usages et pluralisme de l’information en ligne’ [2017] Réseaux 9


<http://unesdoc.unesco.org/images/0015/001584/158450e.pdf> accessed 10 April 2018


——, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Crown 2016)


——, ‘Restoring Transparency to Automated Authority’ (2011) 9 Journal on Telecommunications & High Technology Law 235


Pekka Hallberg and Janne Virkkunen, *Freedom of Speech and Information in Global Perspective* (Springer 2017)


Puddephatt A, ‘Right to Information: Practical Guidance Note’

Reisman D and others, ‘ALGORITHMIC IMPACT ASSESSMENTS’: 22


Reviglio U, ‘Serendipity by Design? How to Turn from Diversity Exposure to Diversity Experience to Face Filter Bubbles in Social Media’ in Ioannis Kompatsiaris and others (eds), Internet Science - 4th International Conference, INSCI 2017, Thessaloniki, Greece, November 22-24, 2017, Proceedings (Springer 2017)
https://doi.org/10.1007/978-3-319-70284-1_22


Södersten A, A Companion to European Union Law and International Law (John Wiley & Sons 2016)


Squires CR, ‘Rethinking the Black Public Sphere: An Alternative Vocabulary for Multiple Public Spheres’ (2002) 12 Communication Theory 446


Sweeney L, ‘Discrimination in Online Ad Delivery’ (2013) 56 Communications of the ACM 44

http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html accessed 13 May 2018


Tufekci, Z. and others, ‘The Ethics of Algorithms: From Radical Content to Self-Driving Cars (Background Paper)’ (Centre for Internet & Human Rights, European University Viadrina 2015)


——, ‘General Comment 34: Freedom of Opinion and Expression (Article 19) (2011) UN Doc CCPR/C/GC/34’

‘UNGA Res 59 (1) Calling of an International Conference on Freedom of Information (14 December 1946)’


Voyenne B, Le droit à l’information (Aubier Montaigne 1970)


programmes/Le-pluralisme-politique-et-les-campagnes-electorales/Les-campagnes-audiovisuelles-officielles> accessed 20 May 2018

‘EdgeRank’ <http://edgerank.net/> accessed 15 March 2018


‘Laws’ (Right2Info.org) <https://www.right2info.org/resources/publications/laws-1/> accessed 16 May 2018


‘Mark Zuckerberg - Continuing Our Focus for 2018 to Make Sure The...’ (Facebook) <https://www.facebook.com/zuck/posts/10104445245963251> accessed 29 January 2018

Martin Siele, ‘New KFCB Licensing Fees That Have Angered Content Creators’ (Kenyans.co.ke, 22 May 2018) <https://www.kenyans.co.ke/news/29611-new-kfcb-licensing-fees-have-angered-content-creators> accessed 22 May 2018


