Pay Attention: a Call to Regulate the Attention Market and Prevent Algorithmic Emotional Governance

Franck Michel, Fabien Gandon

Pay Attention: a Call to Regulate the Attention Market and Prevent Algorithmic Emotional Governance

Franck Michel
University Côte d'Azur, CNRS, Inria
franck.michel@inria.fr

Fabien Gandon
University Côte d'Azur, Inria, CNRS
fabien.gandon@inria.fr

Abstract

Over the last 70 years, we, humans, have created an economic market where attention is being captured and turned into money thanks to advertising. During the last two decades, leveraging research in psychology, sociology, neuroscience and other domains, Web platforms have brought the process of capturing attention to an unprecedented scale. With the initial commonplace goal of making targeted advertising more effective, the generalization of attention-capturing techniques and their use of cognitive biases and emotions have multiple detrimental side effects such as polarizing opinions, spreading false information and threatening public health, economies and democracies. This is clearly a case where the Web is not used for the common good and where, in fact, all its users become a vulnerable population. This paper brings together contributions from a wide range of disciplines to analyze current practices and consequences thereof. Through a set of propositions and principles that could be used do drive further works, it calls for actions against these practices competing to capture our attention on the Web, as it would be unsustainable for a civilization to allow attention to be wasted with impunity on a world-wide scale.

1 An unsustainable attention market

Since the advent of mass market in the 50's, media and advertisement providers have relentlessly tried to figure out effective methods to capture our attention and turn it into revenue. During the last two decades, supported by advances in artificial intelligence (AI), major online social media and Web platforms have brought this process of capturing attention to an unprecedented scale. Based almost exclusively on advertising revenues, their business model consists in providing free services that, in return, collect behavioral traces. This data is then used to maximize the impact of advertisements on users by ❶ ensuring their mental availability at the time of being shown the advertisement, and ❷ ensuring that the message meets their interests, beliefs and moods (i.e. targeted advertising). Based on research in psychology, sociology and neuroscience, several actors including online social media, games and Web platforms have engineered techniques capable of very effectively plundering our "available brain time" [37, 27]. We can distinguish two broad categories of such techniques. Firstly, some techniques are explicitly designed to **leverage cognitive biases** as a means to capture attention. For instance, the *likes* collected after posting content activate the brain's dopaminergic pathways (involved in the reward system) and tap into our need for social approval, giving "bright dings of pseudo-pleasure" [37]; notifications of smartphone applications feed our appetite for novelty and surprise such that it is difficult to resist the urge to check them; the pull-to-refresh mechanism [37], alike slot machines, exploits the variable reward pattern whereby each time we pull down the screen we may get an update or nothing at all; infinite scrolling (of news, posts or videos...) traps us because of our fear of missing out important information (FOMO) to the point that we can hardly break the flow; automatic video chaining replaces a deliberate action to continue watching with a required action to stop watching, and entails a frustrating feeling of incompleteness when stopped; etc. Similarly, some techniques harness dark patterns[1] [26] to manipulate users into taking actions or decisions they wouldn't take otherwise. This is typically the case when one accepts all notifications of an application without really noticing it, while deactivating notifications would require an additional, less intuitive, series of actions.

Secondly, recent advances in machine learning allow the training of **content recommendation algorithms** on massive online behavioral data. These algorithms learn to recommend content that not only captures attention but also increases user engagement[2]. They discover the content's key features that help predict whether such content will effectively attract users' attention, and typically end up selecting content related to conflictuality, fear or sexuality [10]. They also learn to exploit humans' negativity bias [61, 59] and, as a consequence, content conveying high-arousal negative emotions (such as anger, resentment, indignation and disgust) are more likely to be read and eventually shared online than those conveying other emotions [50, 35]. Concerningly enough, false information (a broad term including misinformation and other forms of disin-

[1] The legal definition in California is "A user interface is a dark pattern if the interface has the effect of substantially subverting or impairing user autonomy, decision-making, or choice. A business's intent in designing the interface is not determinative in whether the user interface is a dark pattern, but a factor to be considered." CPRA § 7004 (c)

[2] There are multiple definitions of user engagement. In the context of social media, this typically refers to the fact that a user would interact with a content: e.g. like, comment or repost it. Engagement is usually public in that it leaves public traces on the platform, unlike sheer content consumption that remains private [50].

formation) typically relies on such negative emotions as a trick to foster sharing. Finally, recommendation systems may do all this without it being explicit in the features they select, nor in the succinct feedback that some of them happen to provide[3].

Since the amount of attention available is both limited and precious, it would be unsustainable for a civilization to waste it with impunity for questionable or futile purposes [10]. Today, we might precisely be at that moment: while mental time has become a new oil, we have created an attention economy and subsequent attention markets [28, 29] that, although sustainable from an economic point of view, may be unsustainable from a civilization point of view. From these first references, let us define what the term "attention market" refers to in this article.

Definition 1.1 (Attention market) *Economic environment where businesses compete to capture and retain the resource represented by people's focused mental engagement that we call attention.*

The attention market treats attention as a tradable commodity and involves multiple actors: from producers (the end users whose attention is the resource), to content creators whose work is used to capture the attention, brokers who trade and monetize the attention, and consumers who use it for their purposes such as exposing users to advertisements.

In this article, we propose a discussion aimed at spurring introspection and debate within the computer science community. In line with the Web Science Manifesto [6] calling for interdisciplinary approaches to prepare the future of the Web, we bring together and synthesize the conclusions of more than 70 papers and books from a wide range of disciplines to analyze the practices and drifts of these systems designed to capture attention on a worldwide scale. We make the point that, with the initial commonplace goal of making targeted advertising more effective, the generalization of attention-capturing techniques and their use of negative emotions tends to foster radicalization and polarization, amplify the dissemination of false information, spur the emergence of populism, and eventually put a threat on democracies and human societies in general.

Promoting awareness about these issues, this paper is directly related to UN's Sustainable Development Goal[4] (SDG) 16 "Peace, Justice, and Strong Institutions", since it suggests actions to combat the instrumentalization of negative emotions, the associated false information that mechanically increase the level of anger and resentment among populations, and it promotes "societies that respect the right (...) to freedom of expression, and access to information"[5]. By pointing to the rise of populism worldwide, it addresses the connected question of how to strengthen institutions. The paper is also relevant with respect to SDG 3 "Good Health and Well-being" considering the aggravating effects of online social media on mental health, and the public health issues caused by false information (e.g. during the Covid-19 pandemics).

So far, the public and private research in computer science has invested large efforts in dealing with some aspects of the problem like radicalization, violent speech and false information. These works rely on *post hoc* measures such as content detection, deletion or downgrading. Nevertheless, we argue that additional measures must be considered to actively prevent the issues that stem from attention capturing rather than only mitigating their impact after they have occurred. Presumably, such measures would be political as well as technical, meaning that this socio-technical problematic situation calls for socio-technical solutions. And although the measures may not be associated with immediate research opportunities for the computer science community, we believe that the potential impacts are crucial enough for the community to be fully aware of, and actively involved in, this reflection.

In the rest of this paper we will first review the general principles of recommendation systems and the consequences of the recommendation loop that they implement (section 2). Then, we will explain how having recommendation systems harness emotions can lead to detrimental situations including what we shall name an algorithmic emotional governance (section 3). We will touch upon the threat to creative jobs (section 4) and then review some known post hoc measures (section 5), before discussing preventive measures to reclaim our attention (section 6).

2 Users in the loop... of recommendation systems

Content recommendation algorithms are a key component of a wide range of applications, including social media, search engines and major Web platforms in general. Through many applications they have changed our lives, helping us to be more efficient, assisting us in daily tasks, or improving our education and information. In a number of other applications however, the truth in not so bright. In the case of social media for instance, they are presented to us as if designed to provide us with content that matches our needs and desires, while what they really seek is to maximize the attention we pay to their hosting platform and advertisements thereof.

Through the training process, recommendation algorithms automatically learn to extract from massive behavioral traces the content's features that most effectively capture our attention and maximize our continuous engagement with the platform. For instance, they can learn that some categories of topics, such as conflictuality, fear or sexuality, irresistibly attract our attention [10], and thus lean toward recommending these particular categories. They can also learn to select content tailored for a certain user by taking into account the content's features (topics, source, emotions conveyed...) and its adequacy with the user's profile (interests, inclinations, past behavior...). This adequacy likely involves many other features that are not transparent since the platforms rarely inform users about how and for which purpose their personalized feed was composed. This is underlined in a study by DeVito [18] who analyzed Facebook's patents, press releases, and Securities and Exchange Commission filings, to identify "the set

[3] For instance, a recommendation system may tell us "you liked this movie, you may also like this one". But we don't know what features were selected to recommend this one: Do they have an actor in common? Did my contacts like both of them? etc.

[4] https://www.un.org/sustainabledevelopment/

[5] https://www.un.org/sustainabledevelopment/peace-justice/

of algorithmic values that drive the News Feed". Some of the features he identified are objective, i.e. they can be observed or measured: friend relationships, explicitly expressed user interests, prior engagement, post age and page relationships. By contrast, other features are up to interpretation and thus raise multiple questions: implicitly expressed user preferences (what are the signals of such implicit expression?), platform priorities (what are they and who decides them?), content quality (what are the quality criterion?).

Finally, it may seem that recommendation algorithms learn to leverage psychological traits and cognitive biases. Yet, it is important to stress that the algorithm does not discover such things as a psychological trait or a cognitive bias itself. Rather, it discovers the features that enable it to exploit what psychologists would describe as a trait or bias. Such criteria are not explicitly formulated, they may not even be explicable nor verifiable. They remain implicit in the models unless a study be carried out a posteriori, that would surface the biases that emerge from the recommendations. This is yet another example where AI techniques without explanations nor feedback are problematic.

Another specificity of recommendation algorithms is that they tend to implement a self-reinforcing loop that we define as follows:

Definition 2.1 (Self-reinforcing recommendation loop) *The continuous cycle of recommendation systems providing personalized suggestions to a user based on data collected from their preferences and behaviors and integrating these to further recommendations.*

A classical self-reinforcing recommendation loop is illustrated in figure 1: ❶ The algorithms recommend content to the user. ❷ The behavior of the user is captured, possibly partially due to the focus of the platform and the limited choices that the interface offers, and possibly biased due to the fact that these choices may be oriented, again by the interest of the platform and the chosen interface. ❸ The algorithms integrate these reactions in future recommendations. As a result, the reactions of the user will reinforce the recommendation and propagation of the attached content.

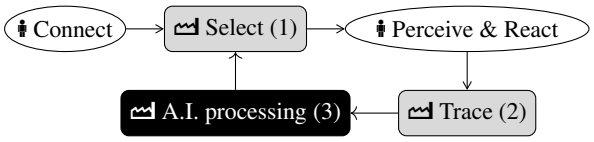

Figure 1: The self-reinforcing recommendation loop of platforms: the ellipses are activities on the user side, the boxes are activities on the platform side. *Select* and *Trace* are grey boxes because only partially observable. The A.I. processing is, more than often, a black box for the end-user.

Of course there are externalities to that loop, that can increase its impact. Smartphones, for example, provide additional means to profile users by tracking their every moves, making recommendation even more efficient and targeted to the point that it competes and sometimes takes over more traditional ways of advertising [65]. Another (detrimental) externality of this loop is that it opens the door to spoofing techniques and other malevolent actors intentionally biasing usage traces to "hijack" recommendation systems. Indeed, as soon as a process is known and documented, it runs the risk of being diverted from its original purpose and manipulated beyond its original objectives. For instance, fake reviews and reactions alter recommendations; black hat techniques of SEO (Search Engine Optimization) such as hidden texts, link farms, cloaking[6] or text spinning are disapproved by search engines as they impact the recommendations they make by unduly increasing the ranking of targeted pages or avoiding their downgrading.

As a result, the fact that a few recommendation systems influence a significant fraction of the human population may have a number of detrimental side effects on their users and our societies at large. A first side effect is that recommendation algorithms tend to lock users in an informational space in accordance with their tastes and beliefs, a "filter bubble" [45] that confines them to a "cognitive comfort zone" and activates their confirmation bias as they are faced with information which seems to go towards the same directions or conclusions [55, 34] Eventually, users are no longer confronted with contradiction, debate nor disturbing facts or ideas, and this algorithmic amplification tends to be a powerful driver of the radicalization and polarization of opinions, leading to extremist ideas in some cases [74].

Furthermore, at a time where we need to change our behaviors (e.g. over-consumption of goods and energy) and redirect our attention to important matters (e.g. climate change), we should question whether recommendation algorithms make the right recommendations, and for whom. Considering the billions of users caught in recommendation loops everyday[7], it is important to continuously monitor how and for what purpose these systems capture our attention. Because when our attention is spent on a content chosen by these platforms, it is lost for anything else.

3 Algorithmic Emotional Governance

Considering the platform's recommendation loop introduced in section 2, we now want to stress that, directly or indirectly, emotions are a key feature of the selected recommendations. In fact, the whole attention market could be seen as driven by a complex equation involving, at least, emotions, cognitive biases and content recommendation algorithms. This could lead to what we will call here an *algorithmic emotional governance* merging two concepts: emotional governance [48] which is the informed management of the emotional dynamics of the governed population, and algorithmic governance [54] which is a governance of our societies based on the algorithmic processing of massive data.

[6] Cloaking denotes a technique in which the content presented to a search engine crawler is different from that presented to an actual user. It aims at deceiving search engines so they display the page that they would otherwise downgrade or dismiss. Adapted from https://en.wikipedia.org/wiki/Cloaking.

[7] In 2018, Google revealed that 70% of the time spent watching videos on Youtube is about videos recommended by Youtube's algorithms. https://qz.com/1178125/youtubes-recommendations-drive-70-of-what-we-watch

Definition 3.1 (Algorithmic emotional governance) *The governance of societies based on algorithms processing massive data to harness the emotional dynamics of the governed population.*

Emotions are a powerful attractor of our attention, especially emotions with a high negative valence [61]. As a result, information that arouses anger, fear, indignation, resentment, frustration or disgust is among those that most effectively catch our attention [50, 35]. An explanation is that witnessing others' negative emotions activates our comparison bias and subjects us to some sort of injunction to take sides, to show our emotional response, and hence publicly demonstrate our "irreproachable morality" [17, 10]. Note that catching attention and increasing user engagement are different things, and although high-arousal negative emotions catch attention more efficiently than other emotions, it remains unclear whether they induce a higher user engagement on social media. In some cases a higher sharing rate of information conveying positive emotions was observed [33, 36]. Nevertheless, in several contributions, researchers showed the overwhelming impact of emotions in argumentation and debates and the means to detect them [5, 68, 4], and it has also been shown that anger spreads faster on social media than any other emotion [21]. Note that this attraction for negative content can be observed in completely different domains, e.g. in literature where the anti-utopian and dystopian fiction genres became more prominent within the utopian genre [39].

Combined together, the construction of filter bubbles by recommendation systems and the ability of these systems to learn the content's features that trigger a particular emotional response in a particular individual, can lead to some form of polarisation and end up trapping users in radicalization pathways. Consider the supporter of a sports club: it is because the system chooses the right topic (e.g. the right sport), the right content (e.g. an article about an opponent club) and the right tone and emotion (e.g. mocking criticism) that an emotion is provoked, followed by a registered reaction (like, comment, repost) and, over time, a potential polarisation is developed such as hatred for the opponent's supporters.

Recommendation after recommendation, the filter bubble becomes an opinion bubble where users are isolated from discrepant opinions, and eventually an emotion bubble where they are maintained in certain emotional states as the result of optimized recommendations. In the end, the complex interaction of negative emotions, cognitive biases (e.g. negativity bias and impulsive tendency to show indignation) and recommendation algorithms leads to an emotional escalation. Often, this escalation is further worsened by the affordances offered by the platforms, that tend to make exchanges ever briefer and more simplistic: How to express a nuanced reflection in a 280-character tweet? How to underline a doubt when the only available choices are essentially limited to ♡/👍 (and sometimes 👎)? How to agree with one part of a post and disagree with another one when this post is treated as a monolithic block by the interface that only offers the options 🔁/↱/🔗 ? This extreme discretization of choices adds to the mechanisms at work and reinforces the polarization of opinions and communities. Some dark patterns are even intentionally employed to make some actions easy and some more difficult: for instance, in Facebook the button to like a post is always visible whereas the option to report a post is at the bottom of a submenu, a pattern falling in the category known as "longer than necessary" [7].

Eventually, nuance, doubt or agnosticism are mechanically made invisible because the low emotional response that they induce simply downgrades their ranking. It is imperative to have an opinion, preferably definite and cleaving. Amplified by digital disinhibition[8] [63], this emotional escalation can lead to outpourings of violence and hatred whose outcome is sometimes tragic as attested by the suicide attempts of teenagers being cyberbullied [56]. Moreover, the full consequences of triggering or regulating emotions on our cognitive functions in general and on memory in particular remain to be studied extensively [49].

We just described the combined effect of emotions, cognitive biases and recommendation algorithms, which is at work whatever the type of content a platform serves. But things get even worse when it comes specifically to false information. False information are frequently meant to arouse strong negative emotions [75], and the combination with cognitive biases and recommendation algorithms provides them with a particularly fertile ground and a formidable cognitive efficiency [2, 38]. Some studies reported that negatively biased fake news enhance users' willingness to share them [14], and reveal a positive correlation between the virality of fake news and the anger they carry [12]. Another study contended that falsehood spreads "significantly farther, faster, deeper, and more broadly than the truth" on social media [69], which underlines that recommendation of content arousing negative emotions does not only induce local individual reaction: it creates a chain reaction leveraging the network effect of social media to spread that "content–emotion" couple through the acquaintance links. Other studies reported that recommendation algorithms mechanically tend to favor false information conveying divisive ideas, shocking events and negative emotions [22, 23]. This type of content entails a felt injunction to take sides and compulsively spread shocking information rather than appealing to critical thinking, questioning its veracity and verifying its source. And since this information is often relayed by acquaintances, the social proof bias [13] entices users to deem it credible and trustworthy.

Concerningly, the contents we are exposed to leave a trace in our implicit memory: although we cannot recall seeing it, it may impact our choices for several months [15]. Even more concerning is the fact that, due to the negativity bias, negative information leaves a longer memory trace than positive information. Therefore, even when a false information is denied or rectified, there remains a negative feeling that stems from the strong emotional response it triggered in the first place. Repeated again and again, associated with representations of the world that summon conspiracy theories, reinforced under the pressure of filter and emotion bubbles, propelled by the network effect, such information gradually and insidiously un-

[8] the feeling of impunity induced by the feeling of anonymity

dermines our trust in the elites (scholars, experts, journalists, politicians, etc.), entails risks for public health [72, 51], and spurs the emergence of extreme ideas and populism that eventually undermine democracies [74, 1, 30, 23], among other pitfalls.

Finally, let us stress that if "previous studies have shown how personality, values, emotions and vulnerability of users affect their likelihood to propagate misinformation" [22], in this section we only considered an average user without any particular health condition. But we should envisage more complex situations when it comes to users with disabilities or mental disorders e.g. depression, anxiety, compulsive shopping disorder, paranoia, FOMO, FOBO[9]... Let us just mention one specific condition: the attention deficit (AD) disorder. There is evidence that AD symptoms could be worsened by the use of digital media and their attention-grabbing applications, and more importantly that these applications could provoke AD among people without previous record of such a disorder [47]. To the very least, more research is needed in this respect.

4 Attention, attention, all thinkers

We firstly intended this section for all the scientists reading this paper, concurring with the article of David R. Smith: "Attention, attention: your most valuable scientific assets are under attack" [60]. In this article, Smith calls for attention to what media platforms are doing to research and the academic domain. Indeed, even the most informed scientists and engineers are not immune to these problems [37] such that digital contraptions (as Smith calls them) are contributing to *academic attention deficit disorder* [60]. In fact, concentration but also boredom, mind-wandering and daydreaming times are vital to creative thinking. Many of us experienced the sudden burst of an idea in the middle of a relaxing moment. Attention-capturing systems steal these moments from all of us and hamper the creativity process of wondering minds [77].

Of course these remarks can be generalized to many other activities requiring concentration, creativity and imagination, and one could wonder what digital contraptions are doing to politics, healthcare or education, for instance. To mention just one example, countless information media now report the cases of Youtubers experiencing a burnout [46], or musicians complaining that they spend more time making Tiktok videos to promote their music than actually creating music [57, 73]. This reveals that, to hook and keep the attention of content consumers, platforms also exercise some sort of visibility tyranny over content creators.

In other words: attention, attention, thinkers, we need to redesign the systems for our own needs, rather than the other way around, especially in creative jobs since the true currency of these jobs are ideas [60].

5 Known Post Hoc Counter-measures

Among the various issues raised in the previous sections, the questions of false information, radicalization, hateful speech and bullying are among the most concerning, and therefore have been extensively addressed by the research community [58]. In [22] authors identified three different points where recommendation systems can be adapted to tackle these issues: ❶ pre recommendation, ❷ within the recommendation model, and ❸ post recommendation. Most of the current counter-measures to deal with false information lie in this third category. Below we touch upon some of them.

Firstly, to dyke the spread of false information as well as inappropriate content such as bullying, hateful or violent speech, social media and content hosting platforms have obligations that vary depending on the legislation and its jurisdiction [24]. Measures range from content deletion and suspension of users spreading inappropriate content, to re-ranking of recommended items before presenting them to the user [22], flagging to indicate potentially deceptive content, etc. Yet, despite these various approaches, progress is still necessary. For instance, subtle violent content may be hard to detect as soon as it does not contain explicit hateful or violent terms, or when it uses sarcasm [44]. Conversely, content may be erroneously assessed as abusive or illicit although it is in fact using irony to convey perfectly acceptable ideas. An in-depth analysis of implicit and subtlety in linguistic content remains an open question [44].

In addition, any action must carefully consider the dangers of transferring regulation and enforcement to private companies. [62] argue that over-filtering content is just as dangerous as letting bad content spread. Indeed, deletion and filtering may deviate from initial purpose to over-censorship of content if it becomes safer for the platforms to do so than take a risk of being sued. Furthermore, assessing the trustworthiness of information raises multiple ethical and political concerns: Who decides what is true or false? According to which criteria? Under whose control?

Secondly, to mitigate the effect of false information, multiple post hoc measures rely on the impact of additional corrective content. For [70], pointing to a coherent alternative explanation, with references to expert and highly credible factual sources, remains a solid starting point. The authors describe the strategy of "observational correction" leveraging the fact that users who witness the correction of a misinformation item, but have not directly engaged with that item, are less affected by cognitive dissonance and are therefore more amenable to correction. This is consistent with the findings of [8] who suggest that exposing users to related stories that correct a post that contained misinformation will significantly reduce misperceptions. The impact of the correction can be further reinforced by explicitly pointing to the demographic similarity between the user and the authors of opposing content [25], which taps into the homophily effect[10]. In other words, we are more likely to accept the correction when it comes from someone who is socially close to us, e.g. having the same professional activity or background. [70] also suggest to multiply correction actions for each targeted content to reinforce the effect.

[9] Fear of Better Options: the inability to choose when faced with a multitude of options.

[10] Homophily: the tendency to associate with similar others.

6 Reclaiming our attention

The methods presented in the previous section all have one thing in common: they deal with the problems in a post hoc manner, that is, after these problems have occurred, with all the limitations that come with this "coming after". To go further however, we need to figure out measures, may they be legal, political or technical, capable of preventing the attention from being looted in the first place. More importantly, we need to consider this reflection not only from the perspective of regulating the attention consumers (the platforms and multiple intermediaries), but also from the perspective of the producers (the end-users) who want to reclaim their attention, especially in times when our attention is needed on a number of urgent matters. This involves actively preventing recommendation systems from finding ways to exploit our inner limitations and manipulate us through sometime ancient and deeply embedded structures of our brain (e.g. our striatum) [9].

Below we formulate a set of propositions stemming from the observations and findings reported in the previous sections. We organize them around the challenges that they address, together with suggestions made by other authors from multiple disciplines. Finally we extract from them a set of empirical principles that could be used do drive further works on good practices.

6.1 The carrot and the stick

Taking the example of false information, [70] insist on the fact that a posteriori corrections are not sufficient and must happen as early as possible, that is, before misperceptions are entrenched. Besides, avoiding the algorithmic amplification effect of such information by recommendation systems requires to mitigate the popularity effect before its happens [22]. But if online social media are required by law to combat false information, they have conflicting incentive to do it, not to say no incentive at all. Indeed, as we described in section 3, false information largely relies on negative emotions to capture users' attention. As such, they are very effective in fostering user engagement which is what online social media strive to obtain. Consequently, from an economic point of view, it is counter-productive for online social media to prevent the spread of false information. More generally, it is counter-productive for platforms to mitigate the popularity effect, mitigate the impact of negative emotions, or reduce filter bubbles and the subsequent polarization of opinions.

The authors of [42] suggest to rethink existing trade regulation laws such as antitrust and fair competition laws under the new realm of attention markets. They propose to enforce taxes on attention consumption to "disincentivize attention intermediaries from vacuuming up as much attention as possible", for instance by restraining the amount of advertisements that can be shown to a user, or reducing the deductibility of advertising expenditures from the companies' revenues to alleviate their taxes. They also propose to regulate the attention costs that can be charged, with the idea that if attention becomes less lucrative then financial resources will be redirected towards more lucrative markets, thus reducing the amount of attention being captured and traded.

In other words, things would not change without strong incentives on one side, and disincentives on the other. We can summarize this in the following general principle:

> 📌 *principle of the right incentive*
>
> At a Web governance level, we must leverage legal and economic means to drive platforms' practices towards desirable behaviours, while penalizing undesired behaviours.

6.2 Usage regulation

Some of the measures meant to regulate the attention market lie in the way the services are consumed. Some legal measures could be taken to enforce a regulation of the daily use of Web platforms. As has already been done in some countries, laws could be voted to limit the daily time spent by users on certain services, especially among the youngest [32, 11]. Another simple measure applies to video streaming platforms, that consists in imposing few-second pauses between videos. This apparently naive technique may actually shatter the infinite feed trap by giving users the short amount of time they need to realize that they have been in an attention tunnel for a while, and that they want to "reclaim" their attention. This can be generalized by formulating the following principle:

> 📌 *principle of supported due diligence*
>
> All means should be provided to foster and update the due diligence of users. In particular they should always be made aware of their options to escape the systems' loops, processes and goals.

Policy makers could also tackle the problem of attention fragmentation entailed by the multiple, often invasive, notifications that smartphone applications raise. Whenever a notification occurs, users are tempted to interrupt their current activity, check the reason of the notification, possibly react to it, before eventually returning to their activity. It has been shown that switching our attention between tasks or contexts has a cost: it is time-consuming and creates a more error prone context [31, 40, 52]. Furthermore it has even been shown that the mere presence of such devices, although turned off, impairs our cognitive capacity [71]. In a way similar to the European General Data Protection Regulation (GDPR)[11] which imposes the consent of users for the use of cookies, law could impose that smartphone applications obtain users' explicit and informed consent for the notifications that they raise, and deactivate them by default ("opt-in only"). Hence the following principle:

> 📌 *principle of opt-in by default*
>
> Recommendation and notification services should be turned off by default and only turned on on demand and after informed consent and preference setting

This could be complemented by more punitive measures, as proposed in [41], for instance by demonetizing and forbidding collaboration with platforms that do not follow the rules.

[11] General Data Protection Regulation (GDPR) https://gdpr-info.eu/

6.3 Content recommendation monitoring

The echo-chamber effect of recommendation algorithms is at the root of multiple examples of polarization and radicalization. It could be mitigated by imposing a certain share of non-recommended content, content that is outside of the user's interests, or content that originates from users they are not acquainted with. In this respect, some approaches lie in the second category proposed by [22], i.e. modifications "within the recommendation" system. For instance, the same authors suggest using clustering approaches to assemble the contacts of the user according to different levels of similarity with the user, and leverage these groups to increase the diversity of recommendations while maintaining a certain coherence and similarity. [20] propose a method to come up with relevant recommendations while reducing the likelihood of enticing the user towards radicalization pathways. Also, to counter the misuse and abuse of anger, indignation or fear, which are often associated with false information, platforms could be required to carry out sentiment analysis on every content in order to keep the amount of recommendations associated to negative emotions below a given threshold.

📌 *principle of balanced recommendations*

Recommendation-based platforms should prevent the over specialization of recommendations w.r.t. all features and should support monitoring and preventing the formation of bubbles of any type (opinion, source, emotion, etc.).

Moreover, there exists an asymmetry of visibility between the viral spreading of an information that was proven to be false or misleading, and the denial or rectification of that information. The denial of a false information usually puts forward a pondered, nuanced position that appeals to reasoning and facts (*logos*) over emotion (*pathos*). Hence, it does not trigger an emotional response compared to the one generated by the false information in the first place, and it is therefore silently downgraded by recommendation algorithms. This is commonly summarized by the so-called Bradolini's law which states that "the amount of energy needed to refute false information is an order of magnitude bigger than that needed to produce it." As a result, users who propagate false information often never get to know about their mistake. [67] insist on the fact that it is critical to jointly address content-checking and digital virality. Thus, to counterbalance this visibility asymmetry, social media could be required to impose on the denial/rectification of a false information a visibility equivalent to that of the initial information, for instance, by ensuring that the population who was exposed to the false information be exposed to the denial too. A warning could also be presented to users who propagated this false information in order to increase their awareness. Of course, this type of measure could be coupled with other post recommendation measures such as strategies involving the observational correction or demographic similarity presented in section 5. More generally, one has to figure out how we can use recommendation systems to recommend counter measures, *i.e.* we could train a recommendation system to learn the most relevant content and the most impactful entries in the acquaintance network to inject a counter measure.

📌 *principle of balanced visibility*

Recommendation-based platforms should ensure that preventive and corrective measures have a visibility at least equal to the visibility of the problems being prevented or corrected.

6.4 Affordances and interaction design

As discussed in section 3, the affordances of platforms are optimized towards extremely brief and basic exchanges, leaving no room for nuance, pondering, doubt nor substantiated reasoning. Interfaces could be redesigned to facilitate non-binary reactions, starting with a range of nuanced emotions. Rather than implementing deceptive dark patterns, they could rely on nudges to gently drive users towards critical thinking, and by valuing/rewarding this kind of behavior. [1] recommend engaging users in the validation of content before sharing it, both manually and with automated analysis methods on content and context. For instance, X (formerly Twitter) asks confirmation before retweeting the link to an article that the user did not click. Similarly, interfaces could encourage users to comment on content instead of merely clicking ♡, 👍 or 👎, and they could question a user about whether they really want to share or support a content associated to strong negative emotions or for which a counter-measure was triggered.

📌 *principle of benevolent interaction design*

Affordances and interactions should be designed and evaluated with the well-being of end-users in mind first.

6.5 Societal impact and educational mission

We, as a society, could decide that large online social media, because of their influence on the society, public opinion, public health and economy, can no longer be considered as sheer private companies regulated by markets law only. Instead, they could be seen as digital commons and be assigned a specific status that would endow them with a societal mission including an educational purpose, for instance. As an example, they could instruct users in detecting false or misleading information, they could promote content meant to increase awareness w.r.t. attention mechanisms and cognitive biases, foster critical thinking and "distill" the scientific method, etc. On the same page, authors of [1] insist on the need for civic education, and [43] recommend integrating democratic values into the algorithms that impact our lives, especially the ones participating in an algorithmic governance (e.g. platform used for debates, for information, for legal actions, educational orientation) which, in our case, means going beyond the optimization of user engagement and attention catching, and including ethical criteria.

📌 *principle of digital commons preservation*

When a digital service, platform or resource reaches the potential of having a world-wide impact on human societies, it must be assigned the status of digital common and must be subjected to preservation rules and policies.

6.6 Feedback and transparency enforcement

Since one of the pitfalls we identified is the fact that users are being caught by the recommendation loops, approaches such as *quantified self* and *lifelog* could be specialized to the case of recommendation-based platforms, in order to foster awareness and introspection. Self-tracking tools could provide users with usage metrics and feedback with respect to the total time spent on the platform, the total exposure to negative news, etc. This could be a way to counter the fact that online sharing of fake news increases with social media fatigue [64]. Indicators could inform users about the diversity of the recommendations they are shown, and make them aware of low-diversity risks. For instance the fact that "90% of the content one sees come from 10% of one's contacts or are on the same topic" may indicate that one is experiencing a filter bubble. We summarize this in the principle below where *user's reflexivity* is the ability of users to be self-aware of their usage and engagement with the system.

📌 *principle of continuous reflexivity*

Users must be provided a continuously updated feedback on their usage of the system and on themselves to support their reflexivity and maintain an up-to-date informed consent.

Among other measures, the European Digital Services Act [66], that took effect in August 2023, requires that platforms set up mechanisms to explain the reasons that led to recommending a certain content, and to offer users an alternative recommendation not based on profiling. Such measures are especially crucial when coupling AI and the Web since we need to set transparency and explanation as a prerequisite to any approach, to ensure the awareness and informed consent of, potentially, billions of users [6].

📌 *principle of full user awareness*

Users must be made aware of all the features and motivations leading to a recommendation, before and when it is provided.

6.7 Build on existing practices

Finally, and although it may seem obvious, one rule is worth remembering: to review and take inspiration from existing best practices in other domains. In most jurisdictions there exist advertising laws to protect consumers, ensure they remain able to make informed decisions, and more generally to maintain a level playing field[12] between all players. Most countries regulate advertising through legislations that target different forms of false, misleading or deceptive advertising contents and claims, and forbid a whole range of practices (unsubstantiated comparison, forged testimonial, puffery, misleading packaging/label, unsolicited commercial messages, alleged contests and sweepstakes, etc.). The work and literature on regulating advertising should be reviewed and built-upon in regulating the attention market at large. This topic is also close to that of clickbaits that are recommended links designed to attract attention and to entice users to follow them while being typically deceptive, sensationalized, or otherwise misleading. Clickbaits are not just teasers but headlines with an element of dishonesty, "using enticements that do not accurately reflect the content being delivered"[13]. As far as we know, there is no regulation of clickbait practices on the Web, although some of these techniques bear similarities with the misleading or deceptive advertisement practices that we just mentioned and that, on the contrary, are regulated.

📌 *principle of best practice transfer*

Methods and tools used to regulate similar situations in relevant domains should be surveyed, benchmarked and systematically considered as input to a Web governance.

To give another example coming from a completely different angle, we know that parenting practices in terms of TV viewing have an impact on the behaviour of young watchers [3]. Again, approaches and good practices in this domain, and more generally in educating and parenting in the digital media age [16], must be considered in the case of "Web viewing" in general and when addressing the problem of attention capturing in particular.

More generally speaking, we need to put in place a governance bodies, starting with the Web and AI, that are prepared to tackle new problematic practices and regulate them, as is done in other areas of activity. And we also need to keep a constant watch on these other areas, if only to draw inspiration from the initiatives and feedback they have on similar issues. Taking the example of the video game industry, there is evidence of a relationship between "loot box"[14] spending and gambling addiction [76], and that a loot box is psychologically akin to gambling [19] and can result in addictive behaviors and endangered players. The way to study and address that unwanted exploitation of our behaviors is inspirational for other problematic practices on the Web such as those we surveyed.

7 Thank you for... your attention

AI is domain-independent. It is being applied in all our areas of interest: information, business, money, politics, employment, sports, games, sex etc. And the worldwide deployment of these techniques, partly due to its coupling with the Web, could have detrimental consequences in all these areas alike, unless properly regulated. This is a commonplace observation but it is the reason why, to prevent such detrimental effects, an ethical AI approach to AI governance must be multidisciplinary and interdisciplinary.

With this mindset, this paper brought together conclusions from more than 70 articles and books from different disciplines (psychology, sociology, neuroscience, politics, legal domain, computer science, education, etc.) to analyze and call for actions against the current practices competing to capture our attention in several "Web Wild West corners". The problem is both critical and complex, and authors of [34] defend

[12] Metaphor denoting the fact that, in business, all players compete fairly, i.e. they all play by the same set of rules. https://en.wikipedia.org/wiki/Level_playing_field

[13] Definition adapted from https://en.wikipedia.org/wiki/Clickbait

[14] "loot boxes" are video game items with randomized contents that can be paid for with real-world money.

the need for a "nuanced multidimensional view of how social media use may shape information consumption" and they urge us to consider "the complex variety in social media platforms [and the] considerable variation in observed impacts among them". In [22] authors add that "This research requires to navigate the careful tension between privacy, security, economic interests, censorship and cultural differences, and requires to be addressed from multiple disciplines that can assess not only the technological aspect, but also the individual and the social one (...) There is ample room for investigation (...), opening a novel, exciting and interdisciplinary line of research."

At the same time, the problem is getting worse with every technological innovation. The pervasiveness of smartphones in our lives has further reinforced the effectiveness of these techniques that can now grab our attention at every moment of the day, and in particular these moments that were previously those of boredom, waiting, daydreaming or intellectual strolling. As we pointed out in section 4, these moments are known to be necessary to spur imagination and creativity. In the continuation of smartphones, smart objects and the resulting internet of things and Web of things will only make things worse.

Recommendation systems that learn to predict us effectively learn to manipulate us, and to be predictable is to lose freedom. Everyday we fuel the predictors in exchange for immediate satisfaction and instant pleasure, this amounts to continually mortgaging our freedom. Besides, these systems that compete for our attention end up pressuring us to consume and to react more and more quickly to their recommendations. And, as we know, acceleration is a form of alienation [53].

In another context and to address our own human limitations, [10] recommended to find ways to increase our overall level of consciousness and reclaim the power of long-term reflection. Our leaders and role models[15] struggle to embody the values of patience, conscience and moderation [10], but our computer systems rarely drive us in that direction either. On the contrary, current AI applications are pushing us not to use our conscience, but to play their automation game. Yet there is no reason for these systems to live in our mind rent free and it is urgent to redesign them so they regularly push us to take a step back, to be more conscious of what we are doing, viewing, saying, spreading, etc. The challenge is to (re)take and (re)give time for awareness, attention and reflection: we need to (re)take that source of freedom. And for this, we proposed a non-exhaustive first set of principles to (re)design Web applications and inscribe in them a set of agreed-upon values.

References

[1] Carlos Abreu and Leandro Ayres França. "Algorithm-driven populism: An introduction [Populizm oparty na algorytmach. Wprowadzenie.]" In: *Archives of Criminology [Archiwum Kryminologii]* (2021).

[2] Alberto Acerbi. "Cognitive attraction and online misinformation". In: *Palgrave Communications* 5.1 (2019), pp. 1–7. ISSN: 2055-1045. DOI: 10.1057/s41599-019-0224-y.

[3] Danielle T. Barradas et al. "Parental Influences on Youth Television Viewing". In: *The Journal of Pediatrics* 151.4 (2007), 369–373.e4. ISSN: 0022-3476. DOI: 10.1016/j.jpeds.2007.04.069.

[4] Valerio Basile et al. "A Pragma-Semantic Analysis of the Emotion/Sentiment Relation in Debates". In: *Proceedings of the 4th International Workshop on Artificial Intelligence and Cognition co-located with the Joint Multi-Conference on Human-Level Artificial Intelligence (HLAI 2016), New York City, NY, USA, July 16-17, 2016*. Vol. 1895. CEUR Workshop Proceedings. CEUR-WS.org, 2016, pp. 117–123.

[5] M. Sahbi Benlamine et al. "Emotions in Argumentation: an Empirical Evaluation". In: *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*. AAAI Press, 2015, pp. 156–163.

[6] Bettina Berendt et al. "Web Futures: Inclusive, Intelligent, Sustainable The 2020 Manifesto for Web Science". In: *Dagstuhl Manifestos* (2021). DOI: 10.4230/DagMan.9.1.1.

[7] European Data Protection Board. *Guidelines 03/2022 on deceptive design patterns in social media platform interfaces: how to recognise and avoid them - v2/0*. Tech. rep. 2022.

[8] Leticia Bode and Emily K Vraga. "That Was Wrong: The Correction of MisinformationThrough Related Stories Functionality in Social Media". In: *Journal of Communication* 65.6 (2015), pp. 619–638. DOI: 10.1111/jcom.12166.

[9] Sébastien Bohler. *Le Bug humain: Pourquoi notre cerveau nous pousse à détruire la planète et comment l'en empêcher*. Robert Laffont, 2019.

[10] Gérald Bronner. *Apocalypse cognitive*. Presses Universitaires de France, 2021. ISBN: 978-2-13-073304-1.

[11] Rory Carroll and Rory Carroll Ireland correspondent. "'Much easier to say no': Irish town unites in smartphone ban for young children". In: *The Guardian* (2023). ISSN: 0261-3077.

[12] Yuwei Chuai and Jichang Zhao. "Anger can make fake news viral online". In: *Frontiers in Physics* 10 (2022), p. 970174. ISSN: 2296-424X. DOI: 10.3389/fphy.2022.970174.

[13] Robert B. Cialdini. *Influence: Science and practice*. Pearson Education, 2008.

[14] Nicoleta Corbu et al. "Fake News Going Viral: The Mediating Effect Of Negative Emotions". In: *Media Literacy and Academic Research* 4.2 (2021), pp. 58–87. ISSN: 2585-8726.

[15] Didier Courbet et al. "The Long-Term Effects of E-Advertising: The Influence of Internet Pop-ups Viewed at a Low Level of Attention in Implicit Memory". In: *Journal of Computer-Mediated Communication* 19.2 (2014), pp. 274–293. DOI: 10.1111/jcc4.12035.

[15] A role model is a person whom others look at as an example to be imitated.

[16] Sarah M. Coyne et al. “Parenting and Digital Media”. In: *Pediatrics* 140.Supplement 2 (2017), S112–S116. ISSN: 0031-4005. DOI: 10.1542/peds.2016-1758N.

[17] M. J. Crockett. “Moral outrage in the digital age”. In: *Nature Human Behaviour* 1.11 (2017), pp. 769–771. ISSN: 2397-3374. DOI: 10.1038/s41562-017-0213-3.

[18] Michael A DeVito. “From editors to algorithms: A values-based approach to understanding story selection in the Facebook news feed”. In: *Digital journalism* 5.6 (2017), pp. 753–773.

[19] Aaron Drummond and James D Sauer. “Video game loot boxes are psychologically akin to gambling”. In: *Nature human behaviour* 2.8 (2018), pp. 530–532.

[20] Francesco Fabbri et al. “Rewiring What-to-Watch-Next Recommendations to Reduce Radicalization Pathways”. In: *Proceedings of the ACM Web Conference 2022*. WWW ’22. New York, NY, USA: Association for Computing Machinery, 2022, pp. 2719–2728. ISBN: 978-1-4503-9096-5. DOI: 10.1145/3485447.3512143.

[21] Rui Fan et al. “Anger Is More Influential than Joy: Sentiment Correlation in Weibo”. In: *PLOS ONE* 9.10 (2014), e110184. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0110184.

[22] Miriam Fernández, Alejandro Bellogín, and Iván Cantador. *Analysing the Effect of Recommendation Algorithms on the Amplification of Misinformation*. 2021. DOI: 10.48550/arXiv.2103.14748.

[23] Piergiuseppe Fortunato and Marco Pecoraro. “Social media, education, and the rise of populist Euroscepticism”. In: *Humanities and Social Sciences Communications* 9.1 (2022), pp. 1–13. ISSN: 2662-9992. DOI: 10.1057/s41599-022-01317-y.

[24] Daniel Funke and Daniela Flamini. *A guide to anti-misinformation actions around the world*. 2023.

[25] R Kelly Garrett, Erik C Nisbet, and Emily K Lynch. “Undermining the corrective effects of media-based political fact checking? The role of contextual cues and naıove theory”. In: *Journal of Communication* 63.4 (2013), pp. 617–637.

[26] Colin M. Gray et al. “Mapping the Landscape of Dark Patterns Scholarship: A Systematic Literature Review”. In: *Designing Interactive Systems Conference*. Pittsburgh PA USA: ACM, 2023, pp. 188–193. ISBN: 978-1-4503-9898-5. DOI: 10.1145/3563703.3596635.

[27] Tristan Harris. *How Technology is Hijacking Your Mind — from a Former Insider*. 2019.

[28] Andreas Hefti and Steve Heinke. “On the economics of superabundant information and scarce attention”. In: *Œconomia. History, Methodology, Philosophy* 5-1 (2015), pp. 37–76. ISSN: 2113-5207.

[29] Vincent F. Hendricks and Mads Vestergaard. “The Attention Economy”. In: *Reality Lost: Markets of Attention, Misinformation and Manipulation*. Cham: Springer International Publishing, 2019, pp. 1–17. ISBN: 978-3-030-00813-0. DOI: 10.1007/978-3-030-00813-0_1.

[30] GlobalData Thematic Intelligence. *Social media, algorithms, and populism*. 2022.

[31] Arthur Thomas Jersild. *Mental set and shift*. 89. Columbia university, 1927.

[32] Alex Kantrowitz. *5 Ways China is Mandating Social Media Changes*. 2021.

[33] Kate Keib et al. “Picture This: The Influence of Emotionally Valenced Images, On Attention, Selection, and Sharing of Social Media News”. In: *Media Psychology* 21.2 (2018), pp. 202–221. ISSN: 1521-3269, 1532-785X. DOI: 10.1080/15213269.2017.1378108.

[34] Brent Kitchens, Steve L. Johnson, and Peter Gray. “Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption”. In: *MIS Quarterly* 44.4 (2020), pp. 1619–1649. ISSN: 02767783, 21629730. DOI: 10.25300/MISQ/2020/16371.

[35] Susann Kohout, Sanne Kruikemeier, and Bert N. Bakker. “May I have your Attention, please? An eye tracking study on emotional social media comments”. In: *Computers in Human Behavior* 139 (2023), p. 107495. ISSN: 0747-5632. DOI: 10.1016/j.chb.2022.107495.

[36] Adam D.I. Kramer. “The spread of emotion via Facebook”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI ’12. New York, NY, USA: Association for Computing Machinery, 2012, pp. 767–770. ISBN: 978-1-4503-1015-4. DOI: 10.1145/2207676.2207787.

[37] Paul Lewis. “’Our minds can be hijacked’: the tech insiders who fear a smartphone dystopia”. In: *The Guardian* (2017). ISSN: 0261-3077.

[38] Cameron Martel, Gordon Pennycook, and David G. Rand. “Reliance on emotion promotes belief in fake news”. In: *Cognitive Research: Principles and Implications* 5.1 (2020), p. 47. ISSN: 2365-7464. DOI: 10.1186/s41235-020-00252-3.

[39] Andrea Burgos Mascarell. “A bibliometric analysis of utopian literature”. In: *ES Review. Spanish Journal of English Studies* 41 (2020), pp. 77–103.

[40] Stephen Monsell. “Task switching”. In: *Trends in cognitive sciences* 7.3 (2003), pp. 134–140.

[41] Cas Mudde and Cristóbal Rovira Kaltwasser. *Populism: A very short introduction*. Oxford University Press, 2017.

[42] John M Newman. “Regulating Attention Markets”. In: *University of Miami Legal Studies Research Paper* (2019). DOI: 10.2139/ssrn.3423487.

[43] Cathy O’neil. *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown, 2017.

[44] Nicolas Ocampo et al. “An In-depth Analysis of Implicit and Subtle Hate Speech Messages”. In: *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*.

Dubrovnik, Croatia: Association for Computational Linguistics, 2023, pp. 1997–2013.

[45] Eli Pariser. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin, 2011.

[46] Simon Parkin. "The YouTube stars heading for burnout: 'The most fun job imaginable became deeply bleak'". In: *The Guardian* (2018). ISSN: 0261-3077.

[47] Chaelin K. Ra et al. "Association of Digital Media Use With Subsequent Symptoms of Attention-Deficit/Hyperactivity Disorder Among Adolescents". In: *JAMA* 320.3 (2018), pp. 255–263. ISSN: 0098-7484. DOI: 10.1001/jama.2018.8931.

[48] Barry Richards. *Emotional governance: Politics, media and terror*. Springer, 2007.

[49] Jane M Richards and James J Gross. "Emotion regulation and memory: the cognitive costs of keeping one's cool." In: *Journal of personality and social psychology* 79.3 (2000), p. 410.

[50] Claire E. Robertson et al. "Negativity drives online news consumption". In: *Nature Human Behaviour* 7.5 (2023), pp. 812–822. ISSN: 2397-3374. DOI: 10.1038/s41562-023-01538-4.

[51] Yasmim Mendes Rocha et al. "The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review". In: *Journal of Public Health* 31.7 (2023), pp. 1007–1016. ISSN: 1613-2238. DOI: 10.1007/s10389-021-01658-z.

[52] Robert D Rogers and Stephen Monsell. "Costs of a predictible switch between simple cognitive tasks". In: *Journal of experimental psychology: General* 124.2 (1995), p. 207.

[53] Hartmut Rosa and Thomas Chaumont. *Aliénation et accélération: vers une théorie critique de la modernité tardive*. La découverte, 2017.

[54] Antoinette Rouvroy and Thomas Berns. "Gouvernementalité algorithmique et perspectives d'émancipation". In: *Réseaux* 177.1 (2013), pp. 163–196.

[55] Kazutoshi Sasahara et al. "Social Influence and Unfollowing Accelerate the Emergence of Echo Chambers". In: *Journal of Computational Social Science* 4.1 (2021), pp. 381–402. ISSN: 2432-2717, 2432-2725. DOI: 10.1007/s42001-020-00084-7.

[56] Ariel Schonfeld et al. "Cyberbullying and Adolescent Suicide". In: *Journal of the American Academy of Psychiatry and the Law Online* (2023). ISSN: 1093-6793. DOI: 10.29158/JAAPL.220078-22.

[57] Neil Shah. "Making TikTok Videos Leaves Musicians Feeling Burnout". In: *Wall Street Journal* (2022). ISSN: 0099-9660.

[58] Karishma Sharma et al. *Combating Fake News: A Survey on Identification and Mitigation Techniques*. 2019. DOI: 10.48550/arXiv.1901.06437.

[59] Michael Siegrist and Gorge Cvetkovich. "Better negative than positive? Evidence of a bias for negative information about possible health dangers". In: *Risk Analysis: An Official Publication of the Society for Risk Analysis* 21.1 (2001), pp. 199–206. ISSN: 0272-4332. DOI: 10.1111/0272-4332.211102.

[60] David R Smith. "Attention, attention: your most valuable scientific assets are under attack". In: *EMBO reports* 19.3 (2018), e45684. ISSN: 1469-221X. DOI: 10.15252/embr.201745684.

[61] Stuart Soroka, Patrick Fournier, and Lilach Nir. "Cross-national evidence of a negativity bias in psychophysiological reactions to news". In: *Proceedings of the National Academy of Sciences* 116.38 (2019), pp. 18888–18892. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1908369116.

[62] Birgit Stark et al. "Are algorithms a threat to democracy? The rise of intermediaries: A challenge for public discourse". In: *Algorithm Watch* 26 (2020).

[63] John Suler. "The Online Disinhibition Effect". In: *CyberPsychology & Behavior* 7 (2004), pp. 321–326. ISSN: 1094-9313. DOI: 10.1089/1094931041291295.

[64] Shalini Talwar et al. "Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior". In: *Journal of Retailing and Consumer Services* 51 (2019), pp. 72–82. ISSN: 09696989. DOI: 10.1016/j.jretconser.2019.05.026.

[65] The Economic Times. "No time to kill: How smartphone is pushing chewing gum out of fashion". In: (2017). ISSN: 0013-0389.

[66] European Union. "Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)". In: *Official Journal of the European Union* (2022).

[67] Tommaso Venturini. "From fake to junk news: The data politics of online virality". In: *Data politics*. Routledge, 2019, pp. 123–144.

[68] Serena Villata et al. "Assessing Persuasion in Argumentation through Emotions and Mental States". In: *Proceedings of the Thirty-First International Florida Artificial Intelligence Research Society Conference, FLAIRS 2018, Melbourne, Florida, USA. May 21-23 2018*. AAAI Press, 2018, pp. 134–139.

[69] Soroush Vosoughi, Deb Roy, and Sinan Aral. "The spread of true and false news online". In: *Science* 359.6380 (2018), pp. 1146–1151. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aap9559.

[70] Emily K Vraga and Leticia Bode. "Correction as a solution for health misinformation on social media". In: *American Journal of Public Health* 110.S3 (2020), S278–S280.

[71] Adrian F Ward et al. "Brain drain: The mere presence of one's own smartphone reduces available cognitive capacity". In: *Journal of the Association for Consumer Research* 2.2 (2017), pp. 140–154.

[72] Przemyslaw M. Waszak, Wioleta Kasprzycka-Waszak, and Alicja Kubanek. “The spread of medical fake news in social media – The pilot quantitative study”. In: *Health Policy and Technology* 7.2 (2018), pp. 115–118. ISSN: 2211-8837. DOI: 10.1016/j.hlpt.2018.03.002.

[73] Dan Whateley. *TikTok’s music influence is ’exhausting’ artists and marketers alike as the industry grapples with the pressure to go viral*. 2023.

[74] Joe Whittaker et al. “Recommender systems and the amplification of extremist content”. In: *Internet Policy Review* 10.2 (2021). ISSN: 2197-6775.

[75] Razieh Nokhbeh Zaeem, Chengjing Li, and K. Suzanne Barber. “On Sentiment of Online Fake News”. In: *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. 2020, pp. 760–767. DOI: 10.1109/ASONAM49781.2020.9381323.

[76] David Zendle and Paul Cairns. “Video game loot boxes are linked to problem gambling: Results of a large-scale survey”. In: *PloS one* 13.11 (2018), e0206767.

[77] Manoush Zomorodi. *Bored and brilliant: How time spent doing nothing changes everything*. Pan Macmillan, 2017.