



Filter Bubbles and the Unfeeling: How AI for Social Media Can Foster Extremism and Polarization

Ermelinda Rodilosso¹

Received: 3 January 2024 / Accepted: 15 May 2024 / Published online: 7 June 2024
© The Author(s) 2024

Abstract

Social media have undoubtedly changed our ways of living. Their presence concerns an increasing number of users (over 4,74 billion) and pervasively expands in the most diverse areas of human life. Marketing, education, news, data, and sociality are just a few of the many areas in which social media play now a central role. Recently, some attention toward the link between social media and political participation has emerged. Works in the field of artificial intelligence have already pointed out that there is a close link between the use of machine learning algorithms in social media and possible epistemic isolation, which could lead to political radicalization. The idea supporting this paper is that artificial intelligence for social media can actively put users' deliberative capacity at risk and foster political extremism. To prove these claims, I proceed along two lines of inquiry. First, I focus on filter bubbles, namely the result of selections made by algorithms that recommend contents that meet users' expectations and opinions. To analyze this phenomenon, I refer to the Deweyan model of experience. Second, I connect the filter bubbles problem to the Deweyan idea of deliberative and participatory democracy and Nussbaum's concept of political compassion. The purpose of this paper is to provide a philosophical foundation that can both (1) effectively serve as a method for analyzing machine learning algorithms and their potential problems in relation to political extremism, and (2) be adopted as a standard to counter the danger of extremism associated with social media experience.

Keywords Artificial intelligence · Machine learning · Ethics · John Dewey · Filter bubbles · Political philosophy

✉ Ermelinda Rodilosso
ermelinda.rodilosso@students.uniroma2.eu

¹ University of Rome Tor Vergata (Department of History, Humanities, Education and Society), Rome, Italy

1 Introduction

The potential of artificial intelligence is multiplying at an exponential rate, to the extent that it is difficult to predict its long-term effects in the social sphere. A platform like ChatGPT is changing the way we engage with intellectual and creative work and, at a more substantive level, the way we search, structure, and learn information (CU Committee, 2023). As stated by an increasing number of studies, generative AI has proven to be effective in the healthcare sector and may be used in areas such as data analysis, medical imaging and clinical diagnosis (ELKarazle et al., 2024; Zhang & Kamel Boulos, 2023; Lang et al., 2024; Teknologirådet, 2022). The algorithms that direct the web and social media are redefining the world of marketing and business, implementing new methods that effectively intercept consumer desires (Poornima et al., 2023; Anayat & Rasool, 2022; Necula, 2023). Deep learning features of AI can significantly improve transport systems and are currently being included in several National AI strategies – Dominican Republic, US, Rwanda, Italy, and India to name a few of them (MEPD, 2023; NSTC, 2023; MINICT, 2022; MID, 2020; NITI Aayog, 2018). These are just some of the many fields in which artificial intelligence is or could be employed. We can therefore say that artificial intelligence constitutes a promising field that is spreading to many areas of human life – and, perhaps, a new phase of technological revolution.

The problem inherent in this technological acceleration is that, since it is progressing very rapidly, we don't have sufficient speculative tools to examine it: simply put, philosophy has not been able to keep up with such a pace. In fact, there seems to be a lack of unitary and cohesive frameworks that can serve as lenses for analyzing issues related to artificial intelligence. Without these theoretical frameworks, the risk is to misuse these tools, either by not grasping their potential or harmfully exploiting them. A clear example of this misuse is the current employment of machine learning algorithms in social media content personalization.

The use of content personalization through machine learning algorithms – which is convenient for marketing reasons and to create engagement between users and platforms – brings with it risks related to the ways in which we acquire information through our social media experience. Indeed, the enormous amount of data we can access is strictly filtered according to our interests and inclinations. It follows that our experience in social media can be so repetitive and self-validating that it promotes the development of epistemic bubbles: we do not have access to all the relevant data we need, but only to the data that adhere to our interests (Leyesen et al., 2022) Anything that does not collimate with those interests is cut off.

Eli Pariser pointed out this issue back in 2011 in the text *The Filter Bubble: What the Internet is Hiding from You*. More than a decade later, the problem is far from resolved. On the contrary, as social media platforms and their functions multiply, the opportunities with which we construct our personal artificial bubbles and remain locked within an epistemic micro-universe also seem to have

multiplied. Such a tendency is particularly problematic in the political field: if we do not reach out to different opinions and perspectives, we expose ourselves to the possible radicalization of our positions and undermining our deliberative capacity. In a nutshell, if we do not have sufficient information about the political environment in which we live, our decisions can be biased and polarized.

The purpose of this paper is to provide a model that can effectively analyze machine learning algorithms and, particularly, their impact on our deliberative and participatory capacity in the political sphere. What is becoming increasingly clear is that the machine learning algorithms currently used to select content in social media tend to discourage the development of critical and pluralistic thinking due to the arbitrary selection of data to which we have access. The idea I intend to propose is that the development of effective philosophical models in the analysis of forms of artificial intelligence can provide concrete resources in dealing with issues related to the application of these tools. In particular, I believe that the Deweyan model of experience can serve as an interpretive grid in content personalization and social media experience and can give practical pointers on how to mitigate the possible harms of Filter Bubbles with special regard to the political field.

2 Selected Hyperconnections

Today it is hard to imagine ourselves disconnected from the digital infosphere and social media. Indeed, they shape our interactions, access to information, and ways of approaching daily life (Petrosyan, 2023a, b; Turkle, 2011). Human existence and the digital environment are so merged that it seems unlikely to conceive them separately. But what does this digital world look like? It consists of hardware, software, networks, infrastructures, and – perhaps most importantly – *data*. The moment we access the Internet, we have at our disposal enormous amounts of data (namely Big Data or Metadata), which are generated on an unprecedented scale and from a multitude of sources (Borgman, 2015; Kitchin & McArdle, 2016). For these reasons, information selection is now crucial for users to employ the Web: without any effective sorting process, we would not be able to navigate a boundless ocean of data. That is why our experience of the digital world is customized and different from anyone else's.

This brings numerous benefits. Users can efficiently access the information and products they are interested in, companies can easily interact with consumers, and the circulation of knowledge has exponentially improved in recent years. At the same time, the selective management of Big Data through specific algorithms gives rise to a paradox: we are both extremely connected with and disconnected from the Web. We constantly interact with the kind of information that is consistent with our habitual behavior on the Web, with little (if any) chance of experiencing unexpected information and data. In other words, we are not guaranteed to be exposed to perspectives other than our own, to explore the diversity of opinions and narratives around us.

It is important to note that exposure to opinions similar to our own and a general narrowness of our experiences also occurs outside the virtual world. Even in the

concrete world, we tend to move within a personal bubble: we hang out with a limited group of people, do a certain type of work, and have access to a specific kind of experience, and therefore tend to receive information and narratives consistent with our daily habits. We have always been dealing with issues related to confirmation bias, selective exposure, and group polarization. In this work, I intend to point out that – considering how selective algorithms work – our social media experience may exacerbate the aforementioned issues.

2.1 A Brief on Recommender Systems (RSs) and Machine Learning Algorithms (MLAs)

To understand how filter bubbles are formed and work – and to better grasp the effects of a preselected informational environment – it is essential to introduce Recommender Systems (RSs) and explain the functioning of Machine Learning Algorithms (MLAs). Recommender systems can be defined as software tools and algorithmic mechanisms that attempt to recommend the most suitable items (e.g., products and services) to a user. They collect and analyze information on the preferences, behavior, and historical data of users to predict their interests in certain items. These recommendations are thus tailored to meet users' interests, needs, and preferences, aiming to facilitate decision-making and increase engagement with items and content (Bobadilla et al., 2013; Ricci et al., 2011).

These systems are undoubtedly useful and can have many applications. First, they allow us to filter the massive amount of information available to users: as I previously said, in today's digital age we tend to be overwhelmed by a massive amount of data, hence recommender systems are crucial for helping users navigate the Web and select the kind of information they need in specific contexts. Second, being exposed to personalized content that adheres to our desires, needs, habits, and dispositions creates a strong bond between users and items. That's why filtering engines are employed in different areas like marketing, education, social life, and entertainment, and thus influence many aspects of our daily lives (Suhaim & Berri, 2021). Given their ability to simplify the users' lives and direct them to content to which they attribute value and significance, these tools have proven to be exceptional resources for fully exploiting the potential of the Web and rapidly became widely used in social media. Moreover, they are becoming much more sophisticated over time. While in first and second-generation social media these engines were more rudimentary, third-generation platforms have seen remarkable advancements in content personalization and have thus created new opportunities and innovations, especially in marketing (Abed et al., 2023).

The point at which recommender systems become crucial in this work lies in the fact that such personalized content is capable of creating engagement with the user (Reviglio & Agosti, 2020). The principle by which machine learning algorithms employed in social media are becoming increasingly sophisticated is that they make the user more involved with the platform. Since users experience a virtual scenario perfectly adherent to their interests, needs, and habits, they will be inclined to spend more and more time on social media to seek gratification. Social media are made to

be appealing and addictive, to intercept the user's attention for as long as possible (Doheny & Lighthall, 2023), and they can achieve that aim through the sense of satisfaction we feel during a social media experience that is consistent with what we consider enjoyable and rewarding (Gal, 2017; Liao & Tyson, 2021; Reviglio & Agosti, 2020). In light of the advantages that a highly interactive user generates, it is crucial to capture their attention with content that they consider interesting, attractive, or content that supports their values and worldviews. Building a comfort zone for the user means more time spent on social media, more efficiency, and therefore more profits (Doheny & Lighthall, 2023).

Hitherto we can infer that recommender systems seem to be very effective tools with which to capture users' interests and direct them to what they consider desirable, and that social media may employ these tools to build an appealing comfort zone for the user. This way, users may be more inclined to constantly engage with the platform. Nevertheless, we are only scratching the surface of the matter. Although people are prone to seek out experiences that conform to what they agree with and find pleasurable (see studies on cognitive dissonance and confirmation bias, e.g. Yahya & Sukmayadi, 2020; Bai et al., 2019), personalization as sophisticated as that generated by machine learning algorithms in social media constitutes a comfort zone with very little permeability to diverse views. It could therefore create a virtual environment prone to polarization and biasing of our opinions (de Arruda et al., 2021; Levy, 2021; Santos et al., 2021; Baldauf et al., 2019; Westerwick et al., 2017) due to the lack of confrontation with opinions and narratives different from ours.

By analyzing the mechanisms of MLAs for content personalization, we get a clearer idea of how accurate the filtering processes of RSs are and how they can result in epistemic bubbles. Schematizing these processes is, however, far from simple. Firstly, social media platforms employ closed source codes (proprietary algorithms), precluding analysis of their specific algorithms. Secondly, the array of algorithms utilized for content personalization is vast, often in combination and continually evolving in sophistication. However, I will try to draw a very simple outline of how they work in broad strokes.

The first stage of MLAs functioning is *data collection*. In this initial phase, engines collect data about users and their interactions with the platform and its content (i.e., demographics, behavior, geography, and psychographics). This data is then separated (*data segmentation*) into smaller parts to extract relevant information and provide more accurate recommendations. Next, we have the *data pre-processing* stage, in which the data are cleaned and rendered in a format suitable for MLAs. *Data analysis* is one of the crucial steps since various algorithms are employed to find meanings and patterns within the data that have been selected. Through this identification of meanings of user-item interactions, preferences, and patterns, it is then possible to figure out how to build a suitable model or framework to optimize the Recommendation System. That is the *model training* phase, where the most suitable algorithms are selected to build an effective predictive model. Subsequently, the *model evaluation* tests and improves the pattern of algorithms and then leads *recommendations generation*, through which users interact with personalized recommendations based on their preferences and interactions. However, personalized

content generation does not exhaust the operation of the algorithms. They are, in fact, trained to monitor and log data to identify new trends, track anomalies and hypothesize possible optimizations (Gangadharan et al., 2024; Farshidi et al., 2023).

This oversimplified summary is useful for us to capture an important point of this analysis. Although our experience on social media appears neutral, accidental, and similar to our daily experiences in the outside world, although what we see in our feed appears to be a serendipitous combination of content and information, it is the result of an ultra-sophisticated system of data selection that aims to analyze and predict what we want to see and experience, excluding from our gaze what the selection deems superfluous or unappealing.

Algorithms are not the result of chance. They are constantly updating and trained to improve themselves *in itinere* and locate any change and anomaly to readjust to a more effective form for the user. Accordingly, our social media experience is not accidental and subject to predetermined mechanisms. The fact that these algorithms work so efficiently in suggesting content that the user considers “comfortable” is the principle on which filter bubble generation relies and what makes these bubbles very little permeable to diversity. As we know, being underexposed to diversity is harmful to democratic life: many works have pointed out how prolonged exposure to similar views in filter bubbles can degenerate into phenomena of polarization and political extremes (Geschke et al., 2019; Mogdil et al., 2021). What I intend to do in this paper is to emphasize how the specific functioning of selective algorithms may foster polarization of opinions and biased decisions and how, as a result, it can impair our deliberative capacity in general. In the next section, I will highlight how the Deweyan theory of experience is useful in understanding the possible consequences of algorithmic selection for our epistemic world.

2.2 Material Experience vs Algorithm-Driven Experience

As mentioned in the previous paragraphs, we are all subject to limited experience. Not only in our digital experience but also in the outside world. As Cass Sunstein argues, “Filtering is inevitable, a fact of life. It is as old as humanity itself. No one can see, hear, or read everything” (Sunstein, 2007: 7). One person cannot access every piece of information or possible experience, and this inevitably makes our views biased in principle, as they are based on a narrow and personal perspective. Moreover, we tend to surround ourselves with like-minded people, we tend to frequent the same environments, and, in general, we are creatures of habit. Epistemic bubbles have always been present in our ordinary lives. Thus, we can conclude that the Filter Bubbles found in social media represent nothing new and pose no additional threat to our ability to form balanced opinions and make informed policy decisions.

However, in the previous paragraph, I showed how the functioning of machine learning algorithms in social media are very sophisticated in suggesting content that systematically conforms to what we already consider valid and positive. Taking this point into consideration, I intend to show how the operation of machine learning algorithms generates a radically different kind of experience compared to the

ordinary one. For this purpose, I refer to the Deweyan theory of experience, which, by showing how experiences influence our knowledge and habits, proves very effective in this kind of analysis.

I would emphasize an important difference between *material experience* – namely, our experience of the tangible world – and *algorithm-driven experience*, which is the kind of experience we have in social media. The former is determined by chance and tends to vary and diversify over time. The latter is pre-selected by an in-depth analysis and constant adjustments to build an a priori network of information and interactions that perfectly adapts to the users' interests. In the first case, experience, though limited, is permeable. If I leave home and follow my usual itinerary to work, I can still interact with experiences that differ from those I previously had. In the second case, experience is not only limited but also impervious. When I open a social network, my experience tends to be identical to the previous ones because it adheres to a certain pattern developed to meet my interests.

In nuce, the presence or absence of the *unexpected* determines a substantial difference between material and algorithm-driven experience. Not only regularities but also chance and serendipitous events direct the material experience of our social and natural environment. Even though we seek out experiences that meet our interests and desires, and we tend to surround ourselves with people that are similar to us, and we follow certain routines and patterns, the unexpected confronts us with kinds of experiences and information that differ from our habits. If I take the same route to work and hang out with the same people for years, I can still encounter deviations from my usual experiences in the material world: being forced to change my route to work because of a construction site, meeting a new person who has no connection to my acquaintances, or acquiring some information unintentionally, are just a few examples of possible variations of everyday life. On the contrary, virtual experience pre-selected by machine learning algorithms tends to be statical, to repeat itself, going back and forth without exploring other kinds of experiential patterns. The data we can access is strictly confined to our pre-existing interests and opinions, making it difficult to go beyond our epistemic world. The fundamental problem with how we experience the virtual world concerns the absence of the unexpected: our virtual experience is entirely predictable and excludes “serendipitous encounters” from our horizon (Sunstein, 2009: 80).

Such predictability has effects on our epistemic world. If we have a much more limited set of experiences, we tend to get accustomed to a form of radical intellectual isolation that has been condensed into the notion of *Filter Bubbles*, namely the result of algorithmic operations that remove from the user's field of vision anything that is considered unsuitable or superfluous to their presumed desires and needs. Eli Pariser was the first to coin the definition, arguing that user experience personalization constitutes a highly problematic aspect of the Web dimension. Essentially developed for commercial purposes, the algorithms behind the filter bubbles create in fact “a unique universe of information for each of us” (Pariser, 2011: 9): when we use search engines, websites, or social media, algorithms prioritize certain data over others based on what kind of information they have about our preferences and our previous searches. According to Pariser, these processes would damage our ability to evolve and learn new ways of thinking and acting, as we struggle to have different

experiences that stimulate our growth. In this regard, he writes “By definition, a world constructed from the familiar is a world in which there’s nothing to learn” (Ivi: 15).

But why is content personalization so harmful for users? Deweyan’s theory of experience may give us some answers. The previous statements contain many similarities with John Dewey’s theory of experience, which Pariser quotes many times throughout *The Filter Bubble: What the Internet is Hiding from You*. I believe that this presence is by no means accidental: Dewey’s theory of knowledge is marked by a naturalistic, scientific, and pragmatist approach, which is well suited to understanding the interaction between humans and artificial intelligence. Unlike other models that reflect on more complex forms of intelligence, his theory – strongly influenced by biology and evolutionism – tends to focus on a basic and generic form of intelligence, and thus, it applies to both human and non-human intelligence. In particular, the Deweyan theory of knowledge has at least three points of intersection with how artificial intelligence relates to humans and their acquisition of knowledge. Artificial intelligence: (a) evolves and refines itself through experience and the accumulation of data; (b) continually adapts to contexts and available data to refine its predictions and responses; and (c) is designed to provide pragmatic answers in real-life situations, through an experimental approach to problem-solving. Experience, adaptation, and experimental approach are three keywords of Dewey’s notion of human and non-human intelligence. In light of these correspondences, I believe that by analyzing the problem of Filter Bubbles through this theory of knowledge, we can find new interpretive and applicative patterns in the field of artificial intelligence and its resonance with human life. My purpose here is to stress the connections between intelligence and experience in relation to the interactions between human and artificial intelligence.

To summarize the Deweyan idea of experience, we can define it as a dynamic *interaction* between an organism and what at that moment constitutes its natural or social environment (Dewey, 1938: 43–4), that both actively engage with different stimuli and reactions. This interaction serves as a foundation for learning, knowledge, and growth and involves elements such as perception, memory, and experimentation (action). As noted by Bernstein and Brodsky, experience affects every aspect of human life, to the point that in Dewey’s conception, there is no dividing line between experience, life, and nature (Bernstein, 1961; Brodsky, 1964; Bernstein, 2010). It is important to say that experience itself is made possible by the phenomenon of *continuity*: the accumulation of certain past experiences will influence future experiences (Dewey, 1938: 37–8). The accumulation of experiences shapes organisms’ responses to their surroundings, and these responses evolve and change according to different contexts and needs. In this way, organisms can learn to better adapt to different types of environments and, as a result, develop new ways of thinking and living. In essence, what enables intelligent organisms to evolve and refine their knowledge is entirely determined by their experiences. Here, I want to underline that the Deweyan idea of experience is based on a direct correspondence between intelligence and experience, between thinking, and learning and conceives intelligence as an ongoing process in which observation, memory, and experimentation-action play a key role.

The same argument applies to human individuals whose intelligence – at a basic level – corresponds to that of any biological being. Therefore, according to what has just been said, human beings grow and evolve through a dynamic interaction with their social and natural environment and through the accumulation of many diverse experiences: the more experiences we have, the better we develop and adapt to our environment, since the fusion of different experiences determines our “capacity to learn” (Alexander, 1987: 130–1; Hutchinson, 2015). If we now return to the previously exposed difference between *material experience* and *algorithm-driven experience*, we can easily grasp how harmful the current system of epistemic isolation generated by filter bubbles can be: an arbitrary and pre-selected limitation, that eliminates any unexpected elements in our daily experiences, hinders our development as human beings and damages our ability to effectively adapt to our environment.

Here it is necessary to refer to an important passage in Dewey’s thought. Although experiences constitute the foundational element of knowledge and human development, Dewey recognizes that not all experiences “are genuinely or equally educative” (Dewey, 1938: 25), that is, they do not all result in ameliorative growth for the individual and may instead result in stunting or degeneration (Pappas, 2014). For this reason, Dewey suggests adopting an experimental attitude toward experiences: these should be continually tested, explored, monitored, and corrected if necessary (Dewey, 1925, 1938). In other words, to learn from our experiences we must analyze them and proceed by trial and error. However, this process, which is fundamental to the development of the human individual, is compromised by filter bubbles because – as expounded in the previous section – the operations of observing, testing, monitoring, and modifying our experiences and information are taken over by machine learning algorithms, which propose to us a pre-selected experience devoid of the experiential elements that, according to Dewey, promote the individual’s development and agency.

In this regard, Mark Coeckelbergh and Elisabeth Anderson point out how filter bubbles undermine subjects’ epistemic agency and can distort and take our opinions to extremes (Anderson, 2021; Coeckelbergh, 2023). I believe that these statements can be further enriched by analyzing more specifically the operation of machine learning algorithms, which shows how our ability to learn from our experiences and develop adequate decision-making power can be hindered by a type of experience that is perfectly tailored for users and, therefore, eliminates the possibility of testing our experiences, and thus our views, opinions, and habits. Based on the above, machine learning algorithms are currently designed in such a way as to: (a) avoid unexpected and serendipitous experiences; (b) hinder the exploration of views and perspectives other than one’s own, making our decision-making more biased and prone to polarization; and (c) prevent human individuals from having an experimental approach to experience. These inherent characteristics of selective algorithms, according to Deweyan experience theory, constitute a detriment to individuals’ growth and the development of their critical sense.

In algorithm-driven experience our thoughts and actions remain confined to our comfort zone and tend to perpetuate themselves consistently. Filter bubbles can thus be considered self-confirming comfort zones that tend to reinforce pre-existing beliefs, behaviors and preferences, while limiting exposure to diverse perspectives

(Courtois et al., 2018; McBride & Amrollahi, 2019). An even more problematic aspect of this phenomenon is that we tend to believe we are all equally connected (or hyperconnected) with the infosphere, when in fact our Web experiences are highly limited and self-referential. In other words, we believe that our way of accessing information is the same as everyone else's and that the entire knowledge and the entire spectrum of virtual interactions are simply there for all to grasp, while we live in a narrow intellectual ecosystem that is mostly impermeable and isolated from other ecosystems.

This isolation can also lead to the generation of echo chambers, namely “the epistemic structure from which other relevant voices have been actively excluded and discredited” (Nguyen, 2020). The process that leads us to favor information that reinforces our pre-existing beliefs (confirmation bias) is a consolidated attitude through which we approach knowledge (Nickerson, 1998). The human mind accumulates and selects information, develops schemata, and tends to reinforce those schemata that appear most credible and functional, to fit more effectively into its ecosystem (Pariser, 2011: 50). However, echo chambers represent a degeneration of confirmation bias and they risk leading the user toward an epistemic selectivity that not only excludes but discredits those who are not part of that sphere of experience and opinion. In this case, users have a clearer level of awareness about the diversity of opinions present outside their bubble but, having found repeated confirmation in their own beliefs, they consider other knowledge fallacious and ineffective.

To summarize, the processes by which algorithms and other forms of artificial intelligence personalize users' algorithm-driven experience: (a) make our algorithm-driven experience monotonous and ineffectual, preventing us from learning and evolving; (b) risk reinforcing in an extreme way our beliefs and opinions, making our deliberative capacity more fallacious and biased. When individuals are constantly exposed to information that aligns with their existing beliefs and preferences, they may become less inclined to critically evaluate alternative viewpoints or engage in meaningful deliberation with others. This can lead to a narrowing of perspectives, a lack of exposure to diverse viewpoints, and a reinforcement of preexisting biases. As a result, deliberative capacity may be diminished, and decision-making processes may become more partisan. From these initial observations, it is possible to see that both filter bubbles and echo chambers strongly alter our experiential environment, and this has important repercussions on the way we perceive and interact with our natural and social world.

3 Democracy as Participation and Deliberation

The harms of filter bubbles are not limited to the cognitive sphere alone. Indeed, what Eli Pariser fears is that the informational determinism that permeates the Web experience can vitiate our civic sense and our ability to make informed political decisions. As mentioned above, our personal bubble filters out data and interactions to offer us only a limited range of possible experiences. In this way, it excludes from its domain a disproportionate amount of data that could enrich the experiential living of the subject. If we return to Dewey's thinking, we can clarify the reasons why

the selection processes from which filter bubbles result can become harmful within the political sphere.

First, it is important to highlight some aspects of the Deweyan conception of democracy. Within Dewey's theoretical framework, democracy does not represent a mere form of government with certain mechanisms and institutional structures. Instead, it is an actual form of life through which human individuals cooperate, form a community, seek a common good, and, at the same time, realize their potential and self-development (Dewey, 1916: 105). Dewey conceives democracy as practical, interactive, and, above all, fundamental to the fulfillment of individuals and their communities. Moreover, it represents the conditions of possibility of social aggregation and constitutes the realization of community life. Such a definition of democracy can be seen as deliberative and participatory: the "free and deliberate participation" (Dewey, 1976: 311) of individuals in social life manifests how human experience condenses into the acquisition of meaning and the achievement of personal and moral growth, from which the individual's community also benefits. As Caspary points out, Dewey's idea of democracy considers public deliberation and participation as an educative experience for all individuals (Caspary, 2000: 10) and hence is a crucial element for human development in general.

However, it is not an effortlessly self-sustaining way of life. On the contrary, Dewey himself recognizes the difficulties associated with maintaining an authentic form of democracy. Indeed, the debate with Lippmann brings out the critical points of the precarious democratic balance, which Dewey acknowledges (Lippmann, 1925; Dewey, 1927). Interestingly, many critical points are related to the ability (or inability) of citizens to acquire a broad level of *information*. While policies should be developed exclusively by experts, the results of their elaboration should be shared with the public, whose acquisition of knowledge should never be manipulated or limited (Dewey, 1927: 224–226) so that anyone can acquire an adequate amount of information on which to base judgements and decisions (Westbrook, 1991: 312). In other words, every citizen must be educated in order to review the information regarding their community and the policies that determine it, as well as to give their assent and actively participate in their implementation. Any democracy that aspires to be both participatory and deliberative must ensure the circulation of information and the exercise of critical analysis.

Needless to say, a condition of experience whose essential feature is the arbitrary selection of information is detrimental to the maintenance of genuine democracy, as it prevents citizens from having an informed opinion on public issues. If we take the Deweyan model of democracy as a point of reference, the selection of data that the public can acquire actually breaks the constructive link between the deliberation-participation binomial and the acquisition of knowledge. Filter bubbles force us inside an intellectual dimension that denies dialogue and critical analysis: if we do not have access to a diverse spectrum of experiences our deliberative capacity falls apart. Beyond that, the individuals fail to reach their full realization and to enhance their capabilities through participation in community life. The meaningful Great Community (Ivi: 170) to which Dewey aspires, a democratic society in which there are no obstacles to human communication and the circulation of knowledge, is reduced to being a kind of ghost dimension marked by solipsism.

Dewey himself identified technological advancement as one of the possible vehicles of social disintegration, or at least one of its concomitant causes. Indeed, *The Public and Its Problems* attributes to technological innovation a remarkable shortening of distances between individuals, which, however, results in a disintegration of face-to-face communities (Ivi: 148–9). The machine age, Dewey writes, “has so enormously expanded, multiplied, intensified and complicated the scope of the indirect consequences, [has] formed such immense and consolidated unions in action on an impersonal rather than a community basis, that the resultant public cannot identify and distinguish itself” (Ivi: 158). In other words, the gradual intensification and expansion of human activities due to technology causes the public dimension to disappear, favoring the creation of closed systems that do not allow citizens to identify with the public. The description of a contradictory reality, hyperconnected but at the same time distant, intensified but at the same time impoverished, appears resonant with that of the Web and social media – so consistent as to make Dewey’s statements somewhat prophetic.

In his later works, Dewey asserts that we should engage with democracy with a truly experimental approach (Dewey, 1989: 273; Hildreth, 2012) made up of discussion, exchange of views, verification, and critical analysis of information. More importantly, it should include a varied set of experiences, sometimes conflicting with each other, so that members of the democratic community can *deliberate*, which means discussing the information at hand and developing plans and alternatives (Dewey, 1927: 70–71); and *participate*, which means taking action to develop certain activities that are consistent with the interests and the needs of the community (Ivi: 175). Deliberation and participation put citizens on the same level, allow majorities and minorities to virtuously communicate with each other, and thus enable the experimentation of decision-making processes in the public sphere. In this way, citizens can exercise their capacities and achieve a fullness of meaning that they could not achieve in an individualistic space.

If we are in a rigidly closed dimension such as filter bubbles, the conditions that – according to Dewey – undermine the very exercise of democracy occur: citizens do not have a broad set of information about the needs and desires of their community; they do not communicate with others; they do not identify with their community; as a result, they tend to be insensitive to the issues of others, and they do not work out new ways of living within the public space. Moreover, citizens risk falling back into a dangerous condition of “political apathy” (Ivi: 164) and a consequent “paralysis” (Ibidem) that freezes action by preventing individuals from interacting with their social and political environment. If “the very heart of political democracy is adjudication of social differences by discussion and exchange of views” then any element that can isolate us from other perspectives is detrimental to the political dimension. Since filter bubbles discourage an exchange of different visions and promote a self-reinforcement of our beliefs, they can deeply undermine the balance of democracy and the very ability of individuals to gain ameliorative experiences, develop critical thinking, and a deliberative capacity free from bias.

An even more alarming fact is that filter bubbles can lead to polarization and radicalization of individuals’ opinions. Since people encounter virtual experiences already adhering to their own inclinations, the presence of a sounding board capable

of amplifying their beliefs may lead them to conceive their view of reality as the only reliable one. In this way, they exclude any possibility of comparison with other perspectives. The relationship between filter bubbles and political radicalization has been the subject of a variety of studies belonging mainly to the field of computer science (Chen & Racz, 2021; Huang et al., 2022; Interian et al., 2021), so I believe that this line of inquiry should be further analyzed through the lens of political philosophy. Here, I propose to recall Dewey's theory of experience as an interpretive model of political radicalization. For instance, if we consider a set of similar experiences repeated over time, according to the Deweyan model of experience we can say that they can stratify and give rise to established practices, habits, and beliefs that the individual can use according to environments, contexts, and needs. Similarly, we can argue that experiential hyperstimulation and constant exposure to repetitive political content can lead to the disproportionate reinforcement of opinions and views – to the point of crystallizing into rigid attitudes and beliefs that refuse to admit other interpretations of the public and its problems. Thus, to break this dangerous cycle of reinforcement, we need to ask ourselves how we can have material and virtual experiences that challenge our perspectives.

4 Political Compassion vs Emotional Anesthetization

The Deweyan concept of political apathy finds resonance in a work by Martha Nussbaum dedicated to the so-called *political emotions* (Ure and Frost, 2014). Before delving into a brief reconnaissance of this analysis, we should take a step back and return to the notion of selective algorithms and filter bubbles. Here, I want to point out the distinctive features of data usually selected by algorithms – which explain why social media platforms employ content personalization. This kind of data can be *appealing*, in the sense that it is perceived as pleasant by the person interacting with them (Pariser, 2011: 68). Or it can be *forceful*, meaning that it takes firm hold of users because of some perceived effectiveness. In other words, we consider these data to be valid and credible – mostly because they confirm our ways of thinking (Ivi: 50–1). These characteristics stem essentially from the fact that we developed selective algorithms primarily for commercial purposes and, therefore, must propose a form of content that the consumers could (or should) consider attractive and compelling.

The appealing condition of the selected information results in the phenomenon through which complex or unpleasant issues – such as Climate Change or the death of migrants in the Mediterranean Sea – are pushed aside in favor of more enjoyable and emotionally superficial experiences. This leads us not only to acquire information in an intermittent and presumably incorrect way but also to develop a form of apathy toward issues that, on the contrary, should set our emotional reactions in motion. As Pariser points out: “In a personalized world, important but complex or unpleasant issues—the rising prison population, for example, or homelessness—are less likely to come to our attention at all” (Ivi: 15).

In opposition to this trend, Martha Nussbaum theorizes that emotions – political compassion in particular – are crucial to the maintenance of an authentically liberal

state. Indeed, Nussbaum identifies a relevant “problem in the history of liberalism” (Nussbaum, 2013: 1), namely that it has not sufficiently addressed the issue of emotions in the political realm. Her idea of democracy postulates that emotions play a key role in ensuring stability in the democratic system and preventing divisions within it (Ivi: 3). Emotions are here essential to stimulate adherence to rules and principles in a non-authoritarian way and invite the cultivation of empathy toward the other. For this reason, the state should stimulate the cultivation of political compassion in citizens through works of art and political rhetoric. According to this account, emotions are pliable elements that individuals can shape with social norms and context. The model Nussbaum takes as a reference for implementing what we might call *emotional education* is that of Greek tragedy, as it stages “shared vulnerabilities” (Ivi: 21) with which the viewer can identify and through which can develop compassion toward other human beings. Art, especially those forms that set narrative processes in motion, can retrace the reasons that lead to a given suffering and make the mind receptive to its identification. The cultivation of compassion motivates citizens to respond effectively to the suffering of others and, therefore, can be an important tool for the perpetuation of social justice.

More generally, as stated by Michael Brady (Brady, 2013) human emotions can have epistemic value since they allow us to evaluate the situations we are experiencing. Through this evaluation, we can grasp the social and natural world around us in a more complex way, identifying the characteristics and processes that led us to have a specific reaction rather than another. Thus, we can argue that emotional cultivation enables us to broaden our intellectual sphere. Indeed, emotions amplify our experience, bring us into contact with other individuals, and stimulate our sense of community and sharing.

In addition, both Martha Nussbaum and Brian Treanor (Treanor, 2014) share the idea that emotions are closely related to the narrative sphere. Narratives prove effective not only in the cultivation of emotions but more specifically in the ability to identify with another individual’s story and to personally experience the situations and emotions they went through. Thus, as well as expanding our cognitive range, emotions can affect our moral dimension and refine how we react to contexts that require the exercise of empathy. Against the risk of solipsistically closing ourselves within our filter bubble, we can thus broaden our cognitive and moral horizons through “many lives, many circumstances, and many places” (Ivi: 125). Narratives, through plots, characters, archetypes, and linguistic devices, are able to elicit images and emotions that suggest how we could/should feel and what we could/should do in a given context – and they do so through an evocative power generated by their normative force.

It should be noted that authoritarian politicians or governments can also exploit emotions and narratives for propaganda purposes. We know that political parties expend considerable resources to orient their constituents and build consensus. In light of this, they do not simply report “bare facts” but build narratives that aim to increase their power of persuasion. Just think, for example, about the expressive devices adopted in recent times by far-right parties: reporting a series of data on illegal immigration or foreign crime may in itself leave one’s audience indifferent. Conversely, associating these (mostly fake) data with narratives about the dangers of

alleged “invasions” and “parasitisms” directs the sentiment of one’s audience in an effective, powerful, and strategically oriented manner. The political debate abounds with similar examples, varying according to the positions adopted by those involved: what they have in common is the constant use of narrative tools endowed with an evocative power. Treanor says:

Bare facts do not have normative force. That gravity accelerates objects on the surface of the Earth at about 9.8 m/sec^2 is a fact. But, even for a physicist, it’s not that fact that structures behavior. Rather, it is experience – both personal and narrativial – that shapes our fear of falling and its consequences [...] “if facts tend not to have sufficient normative force to change belief and behavior, what does? Narrative. (Ivi: 179-80).

Although narratives and emotions have their downside, we can argue that their cultivation can stimulate web users to diversify their virtual experiences and broaden their worldviews, breaking the selection-repetition-reinforcement circuit that can lead to political extremization and polarization. Indeed, new narratives stimulate individuals to step out of their comfort zones and explore modes and conceptions of life that they would not normally have experienced. Thanks also to the close link between narrative and emotion, exposure to new narratives that tell stories set outside our bubble (referring to a different geographical, economic, social, or even ethical-moral location) can stimulate in us the development of mechanisms of empathy and compassion toward people, environments, contexts, and situations that we might not otherwise have experienced.

The theory of political emotions thus shows the importance of being exposed to different narratives and perspectives to mitigate our worldviews and spur us on to a more authentic pursuit of the common good. This is consistent with the Deweyan theory of experience: as mentioned earlier, the exchange of visions that takes place in the participatory and deliberative democratic process allows us to identify common problems through confrontation. We need “shared experiences” and “unchoosen exposures” (Sunstein, 2007: 176) in order to be active and cooperative citizens. Having varied experiences can counterbalance our biased perspective, make us more conscious about common issues, and therefore more prone to have balanced decisions aimed at the good of the community.

How to transpose these considerations to the plane of algorithm-driven experience? How do we get out of our epistemic bubbles while making use of the Internet and social media? How can we ensure that we are exposed to an adequate number of narratives that allow us to construct a reliable and varied idea of the world around us? For example, the possibility of designing social media with open source code was exposed, to foster a genuine exercise of democracy and participation within the Web and for greater transparency about the operation of selective algorithms (Gehl, 2015; Mannell & Smith, 2022). This option is based on the general principles of the Open Source Software movement, which calls for the free circulation of information for the common good and progressive human development (Stewart & Gosain, 2006). One promising proposal is to implement “*serendipity*” within algorithms to enable processing “unexpected and valuable information” and thus stimulate “cognitive diversity, creativity, and innovation” (Reviglio, 2017). I believe that Reviglio’s

proposal may provide greater diversity exposure even to those who do not have prior knowledge of algorithms and thus may be a fruitful direction for the development of algorithms designed to avert the risk of producing filter bubbles.

Making source codes open and implementing more serendipity or randomness within machine learning algorithms are undoubtedly two possibilities to mitigate our epistemic bubbles and allow us to go out and explore the world outside our comfort zones. If “the measure of the worth of any social institution, economic, domestic, political, legal, religious, is its effect in enlarging and improving experience” (Dewey, 1916: 9), then two directions capable of broadening our experiential horizon are an important starting point. However, in this regard, I believe a further step needs to be taken. For machine learning algorithms to authentically promote common welfare and human development rather than pose a threat, they need to be designed through ethical analysis. The Ethical design of algorithms could impart a truly decisive change in the role that algorithms and other forms of artificial intelligence can play within democratic life. It could be an antidote to the degenerations that machine learning algorithms can go to, and transform these mechanisms from a tool of control and homologation to a genuinely democratic tool of liberation.

Many ethical patterns could be applied to algorithms to promote human development instead of discouraging it (Floridi, 2023; Floridi et al., 2018; Mittelstadt et al., 2016). I believe one effective path could be training the algorithms through the virtue ethics model, which has as a foundation the enhancement and flourishing of the human being (Farina et al., 2022). The process of cultivating virtues is based, according to the Neo-Aristotelian view (e.g. McIntyre, 1984), on experience and learning: through the exercise of the virtues in different moral situations (whether real or hypothetical), algorithms can be refined by leading the subject (whether human or artificial) toward excellence. In order to design virtuous algorithms, these must confront different data and experiences, take a cross-cultural approach, and train the creative and narrative ability to imagine different moral scenarios. In doing so, algorithms should be trained to present not only serendipity but also discomfort. Getting out of their comfort zone and learning about even uncomfortable narratives is what enables human individuals to learn about different realities, identify injustices, and promote collective well-being and flourishing.

Exploring diverse narratives and moral situations through constant virtue training by algorithmic systems could significantly decrease the negative repercussions that recommender systems have for users, and thus avoid the formation of Filter Bubbles. Virtuous algorithms could even have an ameliorative effect on the evaluative and moral abilities of individuals, who would be exposed to a much wider and different range of realities than their own limited experience, and hence train their *phronesis* or “practical wisdom” (see Aristotle, 2012). The ethical training of algorithms could also have a positive effect on the circulation of information in general, since they could select quality information, excluding fake news and fake data. On the other hand, it should be noted that algorithms could do little to counteract echo chambers: since – unlike filter bubbles – they involve *voluntary* exclusion and discrediting of information by the subject, we could infer that varied exposure to the Web might not be sufficient to counteract this phenomenon. The topic is vast and would require dedicated studies. I suggest, therefore, further investigations into the link

between machine learning algorithms and the political and collective dimensions, and into ethical models that could effectively be applied to the development of *ethical algorithms*.

Philosophical and ethical analysis of artificial intelligence should be encouraged since, as I said at the beginning of this paper, we do not yet have enough speculative tools to keep up with technological acceleration. The digital revolution holds enormous benefits but also introduces unpredictable consequences. That is why we should conceive philosophy not as Minerva's noctule appearing only at the end of history but as a watchful sentinel conscious of history and ready to intercept the changes and needs of our time, prepared to ask questions and provide answers.

5 Conclusion

In this work, I first tried to reconnect the concept of filter bubbles to the Deweyan theory of experience. At a basic level, Dewey's theoretical framework shows us that we better function as human beings if we accumulate plenty of experiences through which we learn different ways to interact with our environment, which may stratify into habits and beliefs. These processes also show how experience is fundamental to the maintenance of democratic equilibrium, as "everything which bars freedom and fullness of communication sets up barriers that divide human beings into sets and cliques, into antagonistic sects and factions, and thereby undermines the democratic way of life" (Dewey, 1988: 227–8; this same quote is given in Pariser's work).

However, if filter bubbles aprioristically select our experiences and propose to us the same content, our intellectual ecosystem will be limited, and – at the same time – it will limit our development and the stability of the social community in which we interact. These processes could jeopardize our deliberative and participatory capacity as they would (a) limit users' access to information, whereby the ability to engage in informed political discussion and unbiased decision-making would be compromised; (b) discourage social aggregation, preventing the development of forms of collaboration that enable citizens to realize their political potential and positively influence their social environment. More alarmingly, the phenomena of isolation and self-reinforcement can lead to radicalization and polarization of the political views of users exposed to filtering algorithms. As we know, radicalization and polarization are not only a threat to democracy but to all forms of civil coexistence.

Nonetheless, I believe it is possible to mitigate phenomena related to cognitive and moral isolation caused by artificial intelligence and, perhaps, direct these instruments toward greater human empowerment and cooperation. If we develop new modes of experience and expand our sympathetic faculties, we can dismantle our bubble or make it more permeable to external stimuli. Among the ways to deal with this purpose, emotional cultivation can predispose us to openness and cooperation with others. Especially if supported by narrative exercise, the cultivation of emotion can show us that human nature is marked by shared forms of vulnerability. By recognizing ourselves in others, by admitting our mutual need for compassion and care, we broaden our epistemic and moral experience, break the vicious circle of solipsism, and achieve our flourishing and harmonious cohesion with our environment.

Authors' Contributions Not applicable.

Funding Open access funding provided by Università degli Studi di Roma Tor Vergata within the CRUI-CARE Agreement.

Data Availability Not applicable.

Declarations

Ethics Approval and Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abed, L. Y., Hamad, M., & Aljaaf, A. J. (2023). A review of marketing recommendation systems. *Al-Kadhum 2nd International Conference on Modern Applications of Information and Communication Technology*, 2591(1). <https://doi.org/10.1063/5.0119651>
- Alexander, T. M. (1987). *John Dewey's theory of art*. State University of New York Press.
- Anayat, S., & Rasool, G. (2022). Artificial Intelligence Marketing (AIM): Connecting-the-dots using bibliometrics. *Journal of Marketing Theory and Practice*, 32(1), 114–135. <https://doi.org/10.1080/10696679.2022.2103435>
- Anderson, E. (2021). Epistemic bubbles and authoritarian politics. In E. Edenberg & M. Hannon (Eds.), *Political epistemology*. Oxford University Press.
- Aristotle. (2012). *Nicomachean ethics* (trans: Bartlett, R.C., & Collins, S.D.). University of Chicago Press.
- Bai, J., et al. (2019). Exploring cognitive dissonance on social media. *IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2019, 143–145. <https://doi.org/10.1109/ISI.2019.8823262>
- Baldauf, J., Ebner, J., & Guhl, J. (Eds.). (2019). *Hate speech and radicalisation online: The OCCI research report*. Institute for Strategic Dialogue (ISD).
- Bernstein, R. J. (1961). John Dewey's metaphysics of experience. *The Journal of Philosophy*, 58(1), 5–14. <https://doi.org/10.2307/2023564>
- Bernstein, R. J. (2010). *The pragmatic turn*. Polity Press.
- Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, 109–132. <https://doi.org/10.1016/j.knosys.2013.03.012>
- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. The MIT Press.
- Brady, M. (2013). *Emotional insight: The epistemic role of emotional experience*. Oxford University Press.
- Brodsky, G. M. (1964). Dewey on experience and nature. *The Monist*, 48(3), 366–381. <https://doi.org/10.5840/monist196448322>
- Caspary, W. R. (2000). *Dewey on democracy*. Cornell University Press.

- Chen, M., & Racz, M. (2021). An adversarial model of network disruption: Maximizing disagreement and polarization in social networks. *IEEE Transactions on Network Science and Engineering*, 9, 728–739. <https://doi.org/10.1109/TNSE.2021.3131416>
- Coeckelbergh, M. (2023). Democracy, epistemic agency, and AI: Political epistemology in times of artificial intelligence. *AI and Ethics*, 3, 1341–1350. <https://doi.org/10.1007/s43681-022-00239-4>
- Courtois, C., Slechten, L., & Coenen, L. (2018). Challenging Google Search filter bubbles in social and political information: Disconforming evidence from a digital methods case study. *Telematics and Informatics*, 35(7), 2006–2015. <https://doi.org/10.1016/j.tele.2018.07.004>
- CU Committee. (2023). *Generative artificial intelligence for education and pedagogy*. Cornell University. <https://research-and-innovation.cornell.edu/generative-ai-in-academic-research/>. Accessed 3 Apr 2024
- de Arruda, et al. (2021). Modelling how social network algorithms can influence opinion polarization. *Information Sciences*, 588, 265–78. <https://doi.org/10.1016/j.ins.2021.12.069>
- Dewey, J. (1925). *Experience and nature*. Open Court.
- Dewey, J. (1976). *The collected works of John Dewey. The middle works, 1899–1924* (Vol. 7, pp. 1912–1914). Southern Illinois University Press.
- Dewey, J. (1988). *The collected works of John Dewey. the later works, 1925–1953* (Vol. 14, pp. 1939–1941). Southern Illinois University Press.
- Dewey, J. (1989). *The collected works of John Dewey. The later works, 1925–1953* (Vol. 15, pp. 1942–1948). Southern Illinois University Press.
- Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. Macmillan Publishing.
- Dewey, J. (1927). *The public and its problems*. Ohio University Press.
- Dewey, J. (1938). *Experience and Education*. Touchstone.
- Doheny, M. M., & Lighthall, N. R. (2023). Social cognitive neuroscience in the digital age. *Frontiers in Human Neuroscience*, 17. <https://doi.org/10.3389/fnhum.2023.1168788>
- ELKarazle, K., Raman, V., Then, P., & Chua, C. (2024). How generative AI is transforming medical imaging: A practical guide. Applications of generative AI. In Z. Lyu (Ed.), *Applications of generative AI*. Springer.
- Farina, M., Zhdanov, P., Karimov, A., et al. (2022). AI and society: A virtue ethics approach. *AI & Society*. <https://doi.org/10.1007/s00146-022-01545-5>
- Farshidi, et al. (2023). Understanding user intent modeling for conversational recommender systems: A systematic literature review. *ArXiv*. <https://doi.org/10.21203/rs.3.rs-3238230/v1>
- Floridi, L. (2023). *The ethics of artificial intelligence: Principles, challenges, and opportunities*. Oxford University Press.
- Floridi, L., Cowsls, J., Beltrametti, M., et al. (2018). AI4People – an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds & Machines*, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Gal, M. S. (2017). Algorithmic challenges to autonomous choice. *Michigan Telecommunications and Technology Law Review*, 25(1), 59–104. <https://doi.org/10.2139/ssrn.2971456>
- Gangadharan, K., Malathi, K., Purandaran, A., et al. (2024). From data to decisions: The Transformational power of machine learning in business recommendations. *ArXiv*. <https://doi.org/10.48550/arXiv.2402.0810>
- Gehl, R. W. (2015). The case for alternative social media. *Social Media + Society*, 1(2), 1–13. <https://doi.org/10.1177/2056305115604338>
- Geschke, D., Lorenz, J., & Holtz, P. (2019). The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. *British Journal of Social Psychology*, 58(1), 129–149. <https://doi.org/10.1111/bjso.12286>
- Hildreth, R. (2012). Word and deed: A Deweyan integration of deliberative and participatory democracy. *New Political Science*, 34(3), 295–320. <https://doi.org/10.1080/07393148.2012.703852>
- Huang, Z., Silva, A., & Singh, A. (2022). Pole: Polarized embedding for signed networks. In K. S. Candan, et al. (Eds.), *Proceedings, of the Fifteenth ACM International Conference on Web Search and Data Mining*. Association for Computing Machinery, 390–400. <https://catalog.libraries.psu.edu/catalog/41952652>. Accessed 13 Apr 2024
- Hutchinson, D. (2015). Coming to understand experience: Dewey’s theory of experience and narrative inquiry. *Journal of Thought*, 49, 3–17.

- Interian, R., Moreno, J., & Ribeiro, C. (2021). Polarization reduction by minimum-cardinality edge additions: Complexity and integer programming approaches. *International Transactions in Operational Research*, 28, 1242–1264. <https://doi.org/10.1111/itor.12854>
- Kitchin, R., & McArdle, G. (2016). What makes big data, big data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, 3(1). <https://doi.org/10.1177/2053951716631130>
- Lang, O., et al. (2024). Using generative AI to investigate medical imagery models and datasets. *EBio-medicine*, 102. <https://doi.org/10.1016/j.ebiom.2024.105075>
- Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review*, 111(3), 831–870. <https://doi.org/10.1257/aer.20191777>
- Leysen, J., Michiels, L., Smets, A., & Goethals, B. (2022). What are filter bubbles really? A review of the conceptual and empirical work. In A. Bellogin, et al. (Eds.), *Adjunct proceedings of the 30th ACM conference on user modeling, adaptation and personalization*. Association for Computing Machinery.
- Liao, T., & Tyson, O. (2021). “Crystal is creepy, but cool”: Mapping folk theories and responses to automated personality recognition algorithms. *Social Media + Society*, 7(2). <https://doi.org/10.1177/20563051211010170>
- Lippmann, W. (1925). *The phantom public*. Mcmillan & Co.
- Mannell, K., & Smith, E. T. (2022). Alternative social media and the complexities of a more participatory culture: A view from Scuttlebutt. *Social Media + Society*, 8(3). <https://doi.org/10.1177/20563051221122448>
- McBride, N., & Amrollahi A. (2019). How to burst the bubble in social networks? *UK Academy for Information Systems Conference Proceedings 2019*, 44. <https://aisel.aisnet.org/ukais201944>
- McIntyre, A. (1984). *After virtue: A study in moral theory*. University of Notre Dame Press.
- Ministerio de Economía, Planificación y Desarrollo (MEPD), República Dominicana. (2023). *ENIA – Estrategia Nacional de Inteligencia Artificial de la República Dominicana*. https://agendadigital.gob.do/wp-content/uploads/2023/10/Final_ENIA-Estrategia-Nacional-de-Inteligencia-Artificial-de-la-Republica-Dominicana-.pdf. Accessed 3 Apr 2024
- Ministero dell’Innovazione e della Digitalizzazione. (MID), Repubblica Italiana. (2020). *Strategia Nazionale per l’Intelligenza Artificiale*. https://www.mimit.gov.it/images/stories/documenti/Strategia_Nazionale_AI_2020.pdf. Accessed 3 Apr 2024
- Ministry of ICT and Innovation (MINICT), Republic of Rwanda. (2022). *Rwanda’s National Artificial Intelligence Policy Framework*. <https://www.minict.gov.rw/index.php?eID=dumpFile&t=f&f=67550&token=6195a53203e197efa47592f40ff4aaf24579640e>. Accessed 3 Apr 2024
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L., (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>
- Mogdil, S., Singh, R., Gupta, S., & Dennehy, D. (2021). A confirmation bias view on social media induced polarisation during Covid-19. *Information Systems Frontiers*, 1–25. <https://doi.org/10.1007/s10796-021-10222-9>
- National Institution for Transforming India (NITI Aayog), India. (2018). *National strategy for artificial intelligence*. <https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf>. Accessed 3 Apr 2024
- National Science and Technology Council (NSTC), United States of America. (2023). *National Artificial Intelligence Research and Development Strategic Plan 2023 Update*. <https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf>. Accessed 3 Apr 2024
- Necula, S. (2023). Exploring the impact of time spent reading product information on e-commerce websites: A machine learning approach to analyze consumer behavior. *Behavioral Sciences*, 13(6), 439–460. <https://doi.org/10.3390/bs13060439>
- Nguyen, C. (2020). Echo chambers and epistemic bubbles. *Episteme*, 17(2), 141–161. <https://doi.org/10.1017/epi.2018.32>
- Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 175(2), 175–220. <https://doi.org/10.1037/1089-2680.2.2.175>
- Nussbaum, M. (2013). *Political emotions*. Harvard University Press.
- Pappas, G. (2014). What difference can “Experience” make to pragmatism? *Symposia. Language or Experience*. <https://doi.org/10.4000/ejppap.322>
- Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. Penguin.
- Petrosyan, A. (2023a). Internet usage worldwide – statistics & facts. *Statista*. <https://www.statista.com/topics/1145/internet-usage-worldwide/#topicOverview>. Accessed 17 Apr 2024

- Petrosyan, A. (2023b). Number of internet and social media users worldwide as of April 2023. *Statista*. <https://www.statista.com/statistics/617136/digital-population-worldwide/>. Accessed 17 Apr 2024
- Poornima, N., et al. (2023). Recommender systems for personalized business marketing: Employing artificial intelligence and business intelligence in machine learning techniques. *Power Engineering and Intelligent Systems*, 325–35. <https://doi.org/10.1007/978-981-99-7216-6>
- Reviglio, U. (2017). Serendipity by design? How to turn from diversity exposure to diversity experience to face filter bubbles in social media. *INSCI 2017: Internet Science*, 281–300. https://doi.org/10.1007/978-3-319-70284-1_22
- Reviglio, U., & Agosti, C. (2020). Thinking outside the black-box: The case for “Algorithmic Sovereignty” in social media. *Social Media + Society*, 6(2). <https://doi.org/10.1177/2056305120915613>
- Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2011). *Recommender systems handbook*. Springer.
- Santos, F. P., Lelkes Y., & Levin S. A. (2021). Link recommendation algorithms and dynamics of polarization in online social networks. *Computer Sciences*, 118(50). <https://doi.org/10.1073/pnas.2102141118>
- Stewart, K. J., & Gosain, S. (2006). The impact of ideology on effectiveness in open source software development teams. *MIS Quarterly*, 30, 291–314. <https://doi.org/10.2307/25148732>
- Suhaim, A. B., & Berri, J. (2021). Context-aware recommender systems for social networks: Review, challenges and opportunities. *IEEE Access*, 9, 57440–57463. <https://doi.org/10.1109/ACCESS.2021.3072165>
- Sunstein, C. R. (2007). *Republic.com 2.0*. Princeton University Press.
- Sunstein, C. R. (2009). *Going to extremes: How like minds unite and divide*. Oxford University Press.
- Teknologirådet. (2022). *Artificial Intelligence in the clinic – six trends for the health service of the future*. Teknologirådet. <https://media.wpd.digital/teknologiradet/uploads/2023/01/Artificial-Intelligence-in-the-Clinic.pdf>. Accessed 3 Apr 2024
- Treanor, B. (2014). *Emplotting virtue: A narrative approach to environmental virtue ethics*. SUNY Press.
- Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.
- Ure, M., & Frost, M. (2014). *The politics of compassion*. Routledge.
- Westbrook, R. B. (1991). *John Dewey and American Democracy*. Cornell University Press.
- Westerwick, A., Johnson, B., & Knobloch-Westerwick, S. (2017). Confirmation biases in selective exposure to political online information: Source bias vs. content bias. *Communication Monographs*, 84(3), 343–364. <https://doi.org/10.1080/03637751.2016.1272761>
- Yahya, A. H., & Sukmayadi, V. (2020). A review of cognitive dissonance theory and its relevance to current social issues. *Mimbar*, 36(2). <https://doi.org/10.29313/mimbar.v36i2.6652>
- Zhang, P., & Kamel Boulos, M. N. (2023). Generative AI in medicine and healthcare: Promises. *Opportunities and Challenges, Future Internet*, 15(9), 286–301. <https://doi.org/10.3390/fi15090286>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.