120499334

What’s in the Black Box?

Introduction

For this project, the initial aim to was to shed light on the ‘black box’. In the technical field, a black box is when a system is viewed primarily by its input and output characteristics. A black box algorithm is one where the user cannot see the inner workings of the algorithm. Although many large platforms openly discuss technical aspects of their algorithms, considerably less have publicised these algorithms’ impacts on users. The focus of this project shifted from discovering what is in the ‘black box’ to how the black box (or suite of black boxes for multi-platform users) impacts users personally and societally. In Dr. Joel Walmsley’s book, ‘Mind and Machine’, the questions ‘can machines think?’ and ‘are we such thinking machines?’ are posed and addressed. The French philosopher, Rene Descartes’, famous first principle ‘Cogito, ergo sum’ or ‘I think, therefore I am’ is valuable in tackling the question of whether machines think, but African philosophy provides a different lens when exploring social networking sites and their users, as Birhane (2021) describes:

> "Relational frameworks emphasize the primacy of relations and dependencies. These accounts take their starting point in reciprocal co-relations. Kyselo, for example, contends that the self is social through and through - it is co-generated in interactions and relations with others. We achieve and sustain ourselves together with others. Similarly, according to the Sub-Saharan tradition of ubuntu as encapsulated by Mbiti's phrase "I am because we are, and since we are, therefore I am", a person comes into being through the web of relations. In a similar vein, Bakhtin emphasized that nothing is simply itself outside the matrix of relations in which it exists. It is only through an encounter with others that we come to know and appreciate our own perspectives and form coherent image of ourselves as a whole entity. By "looking through the screen of the other's soul", he wrote, "I vivify my exterior." Selfhood and knowledge are evolving and dynamic; the self is never finished - it is an open book."

The self is individual-orientated, social networking sites are the very opposite. Disregarding streaming services such as Netflix and Spotify, users should not be understood on an individual level, but as a member of a network: “There is a Zulu phrase, ‘Umuntu ngumuntu ngabantu’, which means ‘A person is a person through other persons.’ This is a richer and better account, I think, than ‘I think, therefore I am.’” (Birhane, 2017) If a person is a person through other persons, which persons they interact with will shape them, and algorithms have become significant in what people experience online, and who and what they are connected with. In the case of politicians and popstars, it is okay to be stupid if you are well connected (Walmsley, 2016) Can these mathematical processes become a reflection of someone’s identity, or is there potential for users to be changed by their experience with said algorithm(s)?

Can Artificial Neural Networks Sufficiently Refute Free Will?

Darby describes views on free will, determinism and the views (compatibilism and incompatibilism) in which free will and determinism can and cannot co-exist. Darby explains Artificial Neural Networks that are designed on a connectionist model which is “influenced by the architecture of a brain, its network of interconnected neurons, and the concept of recreating this architecture by non-biological means”. The project educates the user through use of a gambling game with an integrated artificial neural network (ANN). The gambling game attempts to predict the users’ next move, and provides the user with a score, rating their predictability. The project aims to deconstruct the intuition of feeling free based on a users’ predictability metric, and to allow users to enjoy the game to the best of their ability without trying to outwit the game. Darby thoroughly investigates appropriate design and deployment, comparing different Integrated Development Environments, outlining qualitative and quantitative data considerations, languages, machine learning libraries, graphical user interface and UX. ANN development and implementation is explained, from a high-level overview of the architecture to activation and error functions. Darby concludes, having identified the characteristics and limitations of ANNs, that they could be used as a hypothetical representation of determinism, for the sake of encouraging users of the gambling game to reflect on perceived impact on free-will. It is concluded that ANNs are not technically sufficient to fulfil the role of determinism in the compatibility/incompatibility debate due to physical and mathematical limitations. The project contributes to the field in that the technical ANN & game development are merged with philosophical thought experiments.

Disentangling the role of algorithm awareness and knowledge in digital inequalities: an empirical validation of an explanatory model.

The paper describes the model of compound and sequential digital exclusions (MCSDE (van Deursen et al, 2017)) assumption that the three levels of digital inequalities are related with sequential paths, which determined the level of digital deprivation among internet users. First-level digital inequalities related to access to both devices amd the internet. Second-level digital inequalities encompass digital and internet skills as well as differences in usage, and third level being related to algorithm awareness and knowledge. The paper maps eleven hypotheses: H1: Ubiquity of internet access positively affects internet skills. H2: Ubiquity of internet access has a positive effect on breadth of internet users. H3: Internet skills have a positive effect on breadth of internet uses. H4: Ubiquity of internet access positively affects algorithm awareness. H5: Ubiquity of internet access positively affects algorithm knowledge. H6: Internet skills have a positive effect on algorithm awareness. H7: Internet skills have a positive effect on algorithm knowledge. H8: Algorithm awareness has a positive effect on algorithm knowledge. H9: Algorithm awareness positively affects breadth of internet uses. H10: Algorithm knowledge positively affects breadth of internet uses. H11: Breadth of internet uses has a positive effect on internet outcomes. The paper poses the research question RQ1: How do the hypothesized relationships in the proposed explanatory model interact with gender, age, education, occupation, and income differences among internet users? Using data from a face-to-face Slovenian Public Opinion Survey, varying methods of analysis were implemented to arrive at the path analysis results. A few of the hypotheses in the model confirmed results of the MCSDE in previous research, but what was new to this research is the inclusion of algorithm awareness and knowledge. Ubiquitous access showed to have a positive effect on algorithm knowledge, but H5 (Ubiquity of internet access positively affects algorithm knowledge) was not supported. Although someone might have a higher awareness of algorithms due to access to more devices, it did not mean they knew more about how algorithms work. H9 was another unsupported hypothesis in that algorithm awareness has no significant effect on breadth of use. The "analyses also found the moderating effect of age, education, and income but not gender or occupation (RQ1). The moderation effects indicate that addressing digital inequalities among disadvantaged groups (older, less educated, lower income) will have a disproportionately greater effect on digital inclusion than among better-off groups" with a one-point internet skill increase improving algorithm knowledge by 0.6 points in below-average income, and 0.3 points among above-average income. The paper mentions many caveats and limitations, but contributes in outlining that internet users should be aware of how content is curated, and how what people see online affects their lives online and offline.

Making curation algorithms apparent: a case study of 'Instawareness' as a means to heighten awareness and understanding of Instagram's algorithm

The paper explores the difficulty in defining and understanding of social network site (SNS) algorithms. Subjects used the cookie extension 'Instawareness' which analysed 50 posts per user, to attempt to gain insight into cognitive media literacy (CML), technical media literacy (TML) and attitudes (ATT). Fouquaert & Mechant mention that the “discrepancy between privacy attitudes and privacy behaviour should not be regarded as a paradox” due to multiple reasons provided by research, such as “perceived benefits of participation on SNS’s seem to outweigh observed risk (Kokolakis, 2017)”. The paper goes on to address the ‘Algorithm Paradox’ of less of a paradox, and more of a ‘complex phenomenon’, due to when users of an SNS understand the effects the algorithms have on their newsfeed, they act accordingly. One research question and three research hypotheses followed: RQ: What is the relation between TML and CML of Instagram’s curation algorithm and what are the consequent attitudes towards Instagram for Flemish adults? H1: Average CML is significantly higher for people using Instawareness compared to those who did not. H2: Average general feelings are significantly higher for people using Instawareness compared to those who did not. H3: Average critical concern is significantly higher for people using Instawareness compared to those who did not. The study design was based around a visual feedback tool which invled firstly a log-in to the user’s Instagram account, and then a revelation of the mechanisms behind the Instagram algorithm. Next, few of the random entrants engaged with Instawareness regularly, and then retook the ‘media questionnaire’. There was a drop-off in engagement in the experiment for the placebo group. The questionnaire was readministered to entrants and altered by adding ‘After using Instawareness…’ before each question. Responses were linked to that of pre-test, and there were no significant differences: this allowed the use of between-subject comparisons. The study involved a group consisting mostly (86%) of highly educated people. The test showed a mean increase in algorithm awareness compared to the placebo group. Regarding general feelings pre-to-post-test, there was a decrease overall.

How using various platforms shapes awareness of algorithms

Abstract: This paper examines how the use of multiple platforms is tied to awareness of algorithms. It builds on the premise that users interact with ecologies or environments of technologies rather than single platforms. The study also supplements work on algorithmic awareness by implementing a mixed-method study to account for how Costa Rican users of Netflix and Spotify understood and related to the algorithms of these platforms. This study combined a survey of 258 participants and 21 semi-structured interviews. Findings demonstrate that multi-platform users were more aware of algorithms and carried out more practical actions to obtain algorithmic recommendations than single-platform users. Although user type did not predict participants’ attitudes towards algorithmic recommendations, higher levels of awareness were associated with more positive attitudes towards algorithms. The study also shows that differences in levels of awareness explained users’ emotional arousal derived from algorithms.

The paper proposed four hypotheses: H1: Multi-platform users will be more aware of algorithms than single-platform users. H2: Multi-platform users will carry out more practical actions to obtain algorithmic recommendations than single-platform users. H3: Multi-platform users will have more positive attitudes towards algorithmic recommendations than single-platform users. H4: Multi-platform users will have higher levels of emotional arousal from algorithms than single-platform users. For the study, participants were divided into single-platform users (regardless of which one, Netflix or Spotify) and multi-platform users. To avoid possible bias toward technical knowledge regarding algorithms, participants with a computer science background or related fields were excluded from the study. The final sample composed of 166 multi-platform users and 92 single-platform users, mostly residing in the Central Valley region of Costa Rica (an area with higher population density and high internet access). Questions were posed in relation to awareness, attitudes and emotional arousal of algorithms. Each set being analysed as well as a scale being developed to measure the gratifications the two platforms provide the users. The study found that H1 was supported, "stress[ing] the importance of analysing users' relationship with algorithms as part of 'polymedia' or broader 'digital environments' (Boczkowski & Mitchelstein, 2021; Madianou & Miller, 20213)". H2 was supported in the study's findings, having outlined the typically younger demographic of multi-platform and algorithmically aware users, engaging in more practices to influence algorithmic recommendations than single-platform users, with user actions also including attempts to 'fight' perceptions of algorithmic 'impositions'. It was found that awareness had more weight regarding positive attitudes for H3 than whether someone is a single-platform user or multi-platform user, though multi-platform users still had more positive attitudes towards algorithmics recommendations than single-platform users. Regarding H4, the study's analysis explained levels of emotional arousal in relation to the connection between user type and awareness: "Users assess the role of algorithms and the relevance of their recommendations through an emotional lens that privileges how either movies or music allow them to nurture or cultivate emotions and moods. The use of various platforms was crucial in this process as it allows people to satisfy affective, cognitive, and integrative needs." I found the study's last example of a participant's experience of algorithms on Spotify particularly interesting; Spotify was said to be able to recognise the user's desire to listen to 'sad' or 'melancholic' music, and the user was pleasantly surprised. At the beginning of the paper, under heading '2.1. Awareness and Active Engagement', other studies are mentioned in their demonstration of 'algorithmic imaginations' and 'folk theories', "Bucher (2017) theorised imaginaries as 'ways of thinking about what algorithms are, what they should be, how they function, and what these imaginations, in turn, make possible". It is similar in ways to folk psychology; "Mostly implicit principles that we (the folk) use to explain the mental states and behaviours of themselves and others in everyday life".

The Artificial Intelligence divide. Who is most vulnerable?

This study surveyed 1088 Dutch citizens, breaking them into five groups: the average users, the expert advocates, the expert sceptics, the unskilled sceptics, and the neutral unskilled. The study aims to (1) identify different user groups in terms of users' level of AI knowledge, skills, and attitudes in the online news and entertainment context through a latent class analysis (LCA) and (2) explore these different user groups' demographic characteristics and identify vulnerable groups. Competent AI users are said to benefit from their understanding of AI recommender systems, proactively influencing content filtering, personalisation, and automated decision-making based on users' data. Users with lower AI competencies not only do not benefit from the ability to cater their experience, but are even said to be more susceptible to harm, resulting from "an interaction between the resources available to individuals and communities and the life challenges they face" (Mechanic & Tanner, 2007:1220) "Such heightened vulnerability increases the risk that such users are influenced, persuaded, or even manipulated by the automated recommendations, possibly leading to issues related to data-driven manipulation, misinformation and disinformation diffusion, and the reinforcement of stereotypes and discrimination (eg. Eubanks, 2017, Hoffman, 2019, Mohamed et al., 2020, Pariser, 2011)". As well as these risks there is also a higher likelihood of vulnerable users forming addiction behaviour that negatively effects their psychological and mental wellbeing when exposed to automated recommendations of entertainment (Tso et al.,2022). Considering the Netherlands has a highly digitalised society and near-universal internet access, the study attempts to answer the research questions: RQ1: Which different groups can be distinguished based on AI knowledge, AI skills, and AI attitudes and (2) what is the prevalence of these groups? RQ2: To what extent can demographic and socioeconomic factors (i.e. gender, age, education) predict respondents' membership to different groups based on AI knowledge, AI skills, and AI attitudes? RQ3: What is the relationship between users' privacy protection skills and their membership to different groups distinguished based on AI knowledge, AI skills, and AI attitudes? A survey was conducted, with varying levels of agreement or disagreement offered to react to statements provided. For AI knowledge, respondents reacted to "Websites and apps for news and entertainment show the same content to everyone", for AI skills, six statements such as "I know where to find the settings to change or turn off personalisation by AI" and for AI attitudes, three statements such as "I like it when online media shows me content that was adjusted to my interests and online behaviour by AI" were provided to be reacted to, and for privacy protection skills, seven items such as "I know how to delete the history of websites that I have visited before" were to be assessed. Respondents were asked to indicate the demographics gender, age and educational level, responses such as "I don't know" and "I'd rather not answer" were coded as missing values. To answer RQ1 a LCA was performed and evaluated using log-likelihood (LL), Bayesian information criterion (BIC), Akaike's information criterion (AIC), Lo-Mendell-Rubin adjusted likelihood ratio test (LRT), and the entropy score was calculated. To answer RQ2 and RQ3, they "conducted a multinomial logistic regression with gender, age, and educational level as predictors, privacy protection skills as covariate, and membership to the user groups as dependent variable." The results showed that 22.2% of the respondents were completely unsatisfied with AI applications in all given situations, and only 4.4% were satisfied with AI on online media, 3% for news and platforms and 8.6% for entertainment. The study concluded with four important insights: 1. Empirical evidence for the presence of the AI divide in the online news and entertainment context was provided. Five user groups were defined. While the AI divide might be dynamic, the two most skilled and sceptical as well as the two vulnerable groups proved relatively stationary. 2. The theoretical assumption that users' perceptions and beliefs can be important factors influencing and driving the AI divide was verified. Attitudes play an important role in the AI divide. 3. In line with previous research on the digital divide (Elena-Bucea et al., 2021; Van Deursen & Van Dijk, 2014): gender, age and education significantly predicted users' membership in different groups. It is expected that the explanatory model used may change with the rapid development of AI technologies, so future studies to examine whether the model will differ in the context of AI divide versus conventional digital divide studies should be conducted. 4. Emphasis on the importance of privacy protection skills. Improving vulnerable users' privacy protection skills can lead to higher AI competence, and vice versa. The study recommends more support for vulnerable users, and for AI designers to develop Explainable AI (XAI) which can increase trust and satisfaction. The study notes that in its limitations, less educated and younger users could be underrepresented in the analysis. Countries that are highly digitalised with universal internet access and composed of citizens with high levels of digital skills could benefit from this study's approach, more studies in various countries are recommended to verify assumptions empirically. Unlike other studies which explored the AI divide in relation to income, occupation, location and ethnicity (Elena-Bucea et al., 2021: Enoch & Soker, 2006) this study focussed on gender, age, and education, so full breadth studies considering all or more of the sociodemographic factors could be run in future.

Privacy

Burgoon (1982) defines privacy based on four dimensions: 1. Informational privacy, which captures the individual control over the processing and transferring of personal information, 2. Social privacy, which captures the dialectic process of regulating proximity and distance towards others, 3. Psychological privacy, which captures the perceived control over emotional and cognitive inputs and outputs, and 4. Physical privacy, which captures the personal freedom from surveillance and unwanted intrusions upon one's territorial space. All four can be applied to use of social networking sites, with the first three being more in relation to how much a user decides to share online, and the latter, considering a user's online presence/persona as their territorial space (albeit digital rather than physical). Behaviours are usually referred to as any observable actions that are taken by individuals, while privacy behaviours are generally referred to as any behaviours that are intended to optimise the relationship with others by either limiting self-disclosure or by withdrawing from interactions with others (e.g. (Taylor & Altman, 1975). This might be the case for most users, some, such as content creators and politicians, focus on virality (more on that later). Privacy concerns have been described as "the desire to keep personal information out of the hands of others" (Buchanan, Paine, Joinson, & Reips, 2007, p. 158). Privacy concerns include user's bank account being compromised, use of online banking, use of real name on social media, sharing family photos, disclosing whereabouts, online identity theft, misuse of personal data and wilful deception in communication processes. In 2006, Barnes described the privacy paradox: "Herein lies the privacy paradox. Adults are concerned about invasion of privacy, while teens freely give up personal information. This occurs because often teens are not aware of the public nature of the Internet". Barnes, in the same year, observed four phenomena of social network site use: 1. The large quantity of information disclosed online. 2. The illusion of privacy on Social Network sites. 3. The discrepancy between context and behaviour (indicating that even when people realise that social network sites are a public realm, they still behave as if it was a private place). 4. The users' poor understanding of data processing actions by online enterprises. In 2023, the Central Statistics Office of Ireland released a statement regarding online cookies and measures taken to minimise their effects: "Cookies are part and parcel of our everyday internet experience that collect personal information about internet users and track their online activity. Less than four in ten (39%) change their settings in their internet browser to prevent or limit cookies, while just 28% of internet users used software to limit the tracking of their movements online." This supports Barnes' fourth phenomenon, although behaviour and attitudes can differ, for example, counterarguments to Barnes' privacy paradox: “"The privacy paradox can be dissolved: The results of our study clearly show that online privacy behaviours are not paradoxical in nature but that they are based on distinct privacy attitudes. The privacy paradox can be considered a relic of the past." (Dienlin & Trepte, 2014). Another phenomenon connected with privacy and the internet is that of privacy fatigue: "The increasing difficulty in managing one's online personal data leads to individuals feeling a loss of control. Additionally, repeated consumer data breaches have given people a sense of futility, ultimately making them weary of having to think about online privacy." (Choi, Park, Jung, 2018), privacy fatigue outlines a sense of helplessness in users; "Privacy fatigue reflects a sense of weariness toward privacy issues, in which individuals believe that there is no effective means of managing their personal information on the Internet." (Acquisti, Friedman, 2006). Following from Central Statistics Office (CSO) of Ireland's statement regarding cookies, in 2023 "Almost two-thirds of internet users blocked advertisers from using their data, [...] but less than one-third limited their tracking online". Could this be due to a lack of awareness surrounding tracking in general versus tracking for advertisement, or due to the inconvenience of ads? Of internet users in 2023: Three quarters of those aged 25 to 34 refused to allow use of their personal data for advertising purposes, compared with half of persons aged 65-74. More than six in ten (61%) restricted access to their geographical location in 2023, while just 43% read privacy policy statements when providing personal information. Less than four in ten (39%) took preventative action by changing the settings in their internet browser to prevent or limit cookies while nearly three in ten (28%) used software to limit cookies. Over six in ten (63%) saw online content which they considered untrue or doubtful. Of these, just short of six in ten (58%) checked its integrity by checking sources of information online or by taking part in online/offline discussions on the content. The relevancy of this data to social networking sites is questionable, in that household consumer behaviour for the same year showed: Approximately 80% of people used the internet for shopping, banking, or booking/ordering services online in 2023. Email remained the most popular internet activity with 93% of internet users who were surveyed in 2023 saying they used email, up from 91% in 2022, and finding information about goods or services was the second most popular internet activity in 2023 at 90%, followed by internet or mobile banking (including PayPal, Revolut, Apple Pay, etc.) at 88%, algorithms implemented by e-commerce and banking might have a greater impact to this dataset, with little information available on social networking site usage available. Like The Netherlands, Ireland is a highly digitalised society, so research on algorithmic awareness and knowledge in Ireland should be conducted.

Filter Bubbles

Eli Pariser published a book in 2011: 'The Filter Bubble, What the Internet is hiding from You'. Pariser describes these filter bubbles in that:

> "Internet filters look at the things you seem to like – actual things you've done, or the things people like you like – and tries to extrapolate. They are prediction engines, constantly creating and refining a theory of who you are and what you'll do and want next. Together, these engines create a unique universe of information for each of us – what I've come to call a filter bubble." (Pariser, 2011: 9)

As Pariser (2011: 125) puts it, you click on a link, which signals an interest in something, which means you're more likely to see articles about that topic in the future, which in turn prime the topic for you. You become trapped in a you loop, and if your identity is misrepresented, strange patterns begin to emerge, like reverb from an amplifier.

> Pariser mentions that filter bubbles may have several outcomes for individuals: narrower self-interest; overconfidence; dramatically increased confirmation bias; decreased curiosity; decreased motivation to learn; fewer surprises; decreased creativity and ability to innovate; decreased ability to explore; decreased diversity of ideas and people; decreased understanding of the world; and a skewed picture of the world. Eventually, “You don’t see the things that don’t interest you at all” (Pariser, 2011: 106) because the filter bubble “will often block out the things in our society that are important but complex or unpleasant. It renders them invisible. And it’s not just the issues that disappear. Increasingly, it’s the whole political process” (Pariser, 2011: 151)

Dahlgren 2021 released a critical view of Pariser’s filter bubbles with nine counterarguments:

- 1. Filter bubbles can be seen at two levels: technological and societal
    - Technological refers to the immediate implications to the content recommended of any single choice made, narrowing content available over time. Similar to a live microphone in too-close proximity to a loud speaker and the resulting ever-increasing volume produced by the pair as a same closed system. "Let us call this a filter bubble at the technological (or individual) level."
    - The societal level refers to the broader causes and consequences for humans

- 2. People often seek supporting information, but seldom avoid challenging information
    - "Selective exposure research has shown that people, on average, prefer supporting information to challenging information [...] which is a typical case of confirmation bias."
    - The latter is more significant in counterarguing filter bubbles:
    - People are said to have two motivations: to seek information (which is moderately strong) and to avoid information (which is a comparatively weak motivation)
        - Garrett, 2009; see also Fischer et al., 2011; Frey, 1986; Munson & Resnick, 2010
    - Individuals with extreme views might be thought to avoid challenging information, but "Bruns suggests, "they must monitor what they hate"

  - (2019a: 97). Meta-analytic findings similarly suggest that the more confident an individual is in their belief, attitude or behaviour, the more exposure they have to challenging information (Hart et al., 2009)"
  - "Relatively few respondents across eight countries reported that they saw news on social networking sites that supports their beliefs or attitudes (Matsa et al., 2018)"
  - Reuters Institute's annual survey of the populations of 36 countries, 40% of SNS users agreed with the statement that they were exposed to news they were not interested in, or news sources they did not typically use, 27% disagreed with the statement (Newman et al., 2017)
  - Facebook users who used it for news consumption during the American presidential election in 2016 were more exposed to news that both challenged and confirmed their beliefs and attitudes, with polarisation declining in one study (Beam et al., 2018)

- 3. A digital choice does not necessarily reveal an individual's true preference
  - "As Pariser puts it, "what the code knows about you constructs your media environment, and your media environment helps shape your future preferences" (Pariser, 2011: 233). Given enough time, Internet services are ultimately expected to become "a perfect reflection of our interests and desires" (Pariser, 2011: 12), which can lead to "information determinism, in which our past clickstreams entirely decide our future" (Pariser, 2011: 135)."
  - "we should not necessarily assume that people have an active agency when they select content, but become passive and malleable when they receive information"
  - Choices and preferences differ, choices are actions, while preferences are states of mind
  - "We (or the algorithm) can directly observe what a person selects, but we can never directly observe what a person prefers" Examples provided are opting for a movie about journalism based on the actors rather than an interest in journalism, an atheist visiting a religious web site in order to find counterarguments in a debate, and choices with greater consequences weakening preferences (watching a sitcom being the immediate preference but choosing to study based on the outcome of a job)
  - People often aim to portray themselves in a socially favourable light, clicking 'like' is not always about supporting the information, but also what that 'like' communicates to others (Hart et al., 2019)
  - People are said to click on news items with many likes, to avoid missing out on popular news

  - "Media companies may also try to push content towards individuals through advertisements (Webster, 2017)."
  - Pariser argues that we need "more data and more programmers to "solve" this problem (Pariser, 2011:118). Here we can object by citing the bias-variance trade-off (i.e., underfitting contra overfitting), which means that an algorithm or model may perform very well when it is trained on one set of data but then perform poorly when applied (generalised) to unseen data. These are well-known limitations of all theoretical or statistical models, and they will never go away unless one makes a custom model for each and every individual (this should not be confused with personalisation algorithms, which are not necessarily tailored for each and every individual- rather, it is often the output of personalisation algorithms that is tailored for each and every individual)."

- 4. People prefer like-minded individuals, but interact with many others too
  - This focuses on people rather than technology, political discussions often have crosscutting remarks and there is more to talk about if there are disagreeing remarks. Heterogeneity increases over time with social media use, "the more we use social networking sites, the greater diversity of the people we talk to (Choi & Lee, 2015; Lee et al., 2014)". In a workplace setting, opposing opinions are common, but people often avoid speaking up or revealing actual beliefs or thoughts to avoid conflict (Cowan & Baldassari, 2018)

- 5. Politics is only a small part of people's lives
  - The filter bubble thesis refers to almost only political discussions and information
  - "we typically interact with about 150 people offline (Hill & Dunbar, 2003), and of those, we tend to discuss politics with only about 9 (Eveland & Hively, 2009)"
  - If only users' reaction to controversial or self-supporting political content is measured, there is a sampling bias and filter bubbles do not account for how everyone interacts with technology and it's effects.
- 6. Different media can fulfil different needs
  - Users' media diet can vary, although fully insulated in a filter bubble on one social networking site, users will often visit mainstream news sites
  - "Those in filter bubbles would no longer find out about major news events at all (because we rarely click 'like' on news about terrorist attacks or natural disasters, which are consequently filtered out from our bubbles),

    - which means that social networking sites, such as Facebook, (Pariser, 2011)"
    - Filter bubble thesis assumes a journalistic lens, while sites like Facebook have a goal of connecting people. Should Facebook be considered a news site and algorithms the editor?
- 7. The United States is an outlier in the world
    - "The filter bubble thesis originates from the US. This is a country that, over time, has had the most significant decline in trust in both the press and political institutions among some 50 countries (Hanitzschet al., 2018), has a weak public service broadcaster (public service broadcasters can have a dampening effect on political polarisation; see Bos et al., 2016; Trilling et al., 2017), and has a two-party system where the Senate and Congress have been heavily polarised since the 1970s (Arceneaux & Johnson, 2015)"
    - More analysis comparing other countries to the US needs to be conducted in relation to polarisation.
    - Multi-party system countries have lower levels of polarisation on Twitter than two-party system countries (Urman, 2019)
    - Contrary to the filter bubble thesis, groups in the US who are least likely to use the internet and social media, had the greatest increase in polarisation (Boxell et al., 2017: 10612)

- 8. Democracy does not require regular input from everyone
    - Although exposure to challenging information is necessary for some normative views on democracy, democracy would not stop working if people were insulated in filter bubbles.
    - Many democracies delegate responsibilities of citizens to representatives
    - "People can still follow democratic processes (vote, raise concerns with representatives, protest, demonstrate, etc.) even if they only consume supporting information"
    - Filter bubbles can be one democratic way to form interest groups that influence societal institutions and the mass media (Athaus, 2006)
- 9. It is not clear what a filter bubble is
    - Definitions include: "personal ecosystem of information that's been catered by these algorithms to who they think you are" (The Daily Dish, 2010: para. 1) and "unique universe of information for each of us" (Pariser, 2011: 9)
    - Nothing inherently wrong with definitions, but lacking in clarity

  - "When filter bubbles are discussed, there is a risk of confusing two arguments: the first strong but also trivial and uninteresting, and the second, weak and speculative but also most interesting. This general phenomenon has been called the Motte and Bailey doctrine (Shackel, 2005)"
  - The Bailey in relation to filter bubbles is political polarisation of technology for democracy
  - The Motte retreats to the fact that two users will see different results when searching for the same information

- A filter bubble is said to be "a misunderstanding especially about the power of technology - and in particular, of algorithms - over human communication" (Bruns, 2019: 93)

Understanding Social Media Recommendation Algorithms.

History

Amazon was the first to roll out a large-scale algorithm in the late 1990s, with Netflix not far behind in 2000. Collaborative filtering was employed by these platforms in the early 2000s, "Customers who bought/watched this also bought/watched that". Unexpected combinations of products were discovered by collaborative filtering, like putting beer next to diapers to cater to frazzled fathers. Netflix now uses a more complex approach, classing viewers by their taste/style, of which there are a couple of thousand. "few less obvious sources, like the smart humour of Master of None and the psychological thrill of Making A Murderer driving people towards the wise-ass private detective. Meanwhile, "shows that expose the dark side of society" were shown to drive viewers to Luke Cage, such as the question of guilt in Amanda Knox and the examination of technology in Black Mirror." (Plummer, 2017)

Steering away from privacy in users, and toward publicity and reach, the nature of content is not always civil: "US politicians learned to be less civil because such posts garnered more attention" - (Jeremy A Frimer et al., 2023).

I have included a few explanatory images from Narayanan's publication on social media recommendation algorithms below:

"Figure 1: The effects of information propogation on platforms emerge through the interaction of design and user behaviour, based on underlying mathematical principles.

Design comprises algorithms, the user interface, and various policies, such as content moderation policies. Platform designers, users, and content creators all adapt to emergent effects."

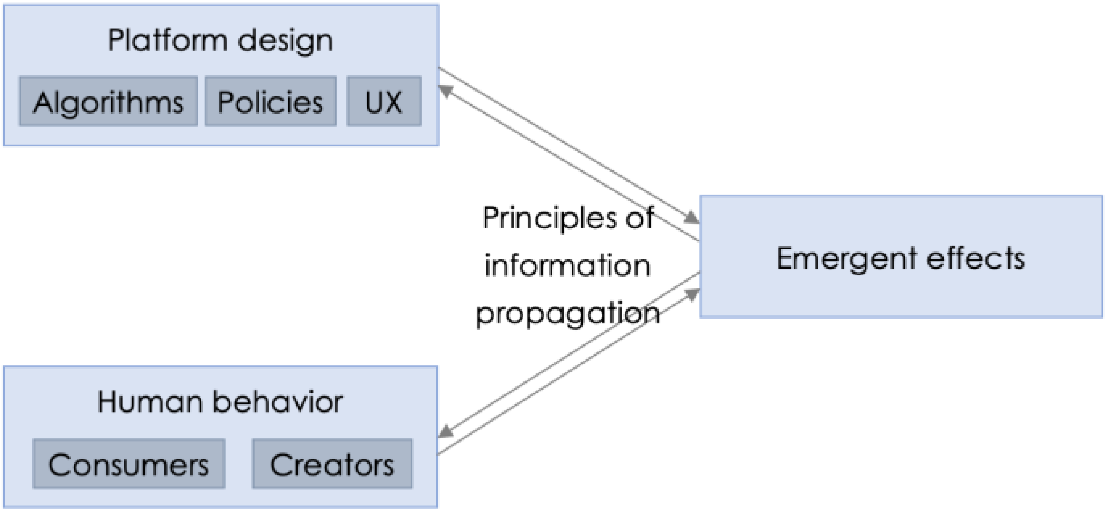
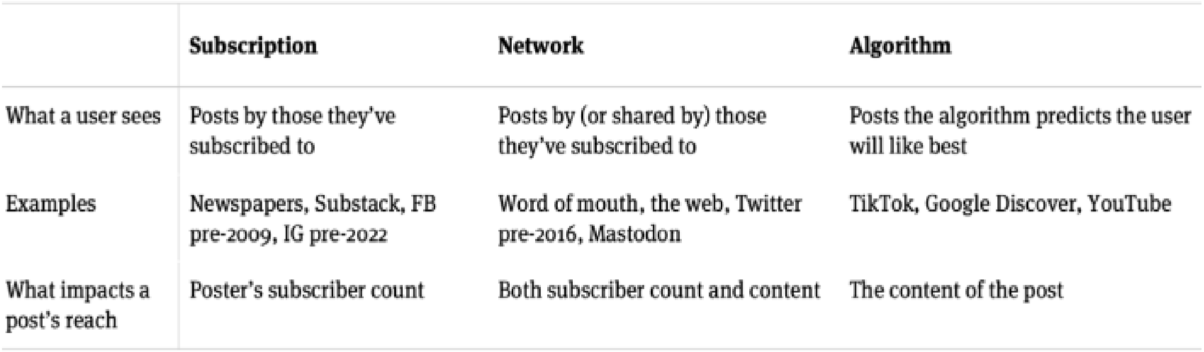
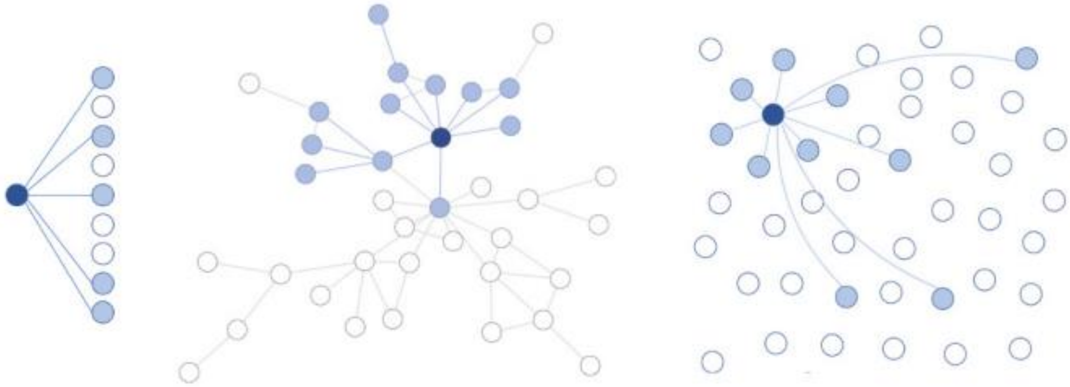


Table 2: Three stylized models of information propagation.



	Subscription	Network	Algorithm
What a user sees	Posts by those they've subscribed to	Posts by (or shared by) those they've subscribed to	Posts the algorithm predicts the user will like best
Examples	Newspapers, Substack, FB pre-2009, IG pre-2022	Word of mouth, the web, Twitter pre-2016, Mastodon	TikTok, Google Discover, YouTube
What impacts a post's reach	Poster's subscriber count	Both subscriber count and content	The content of the post

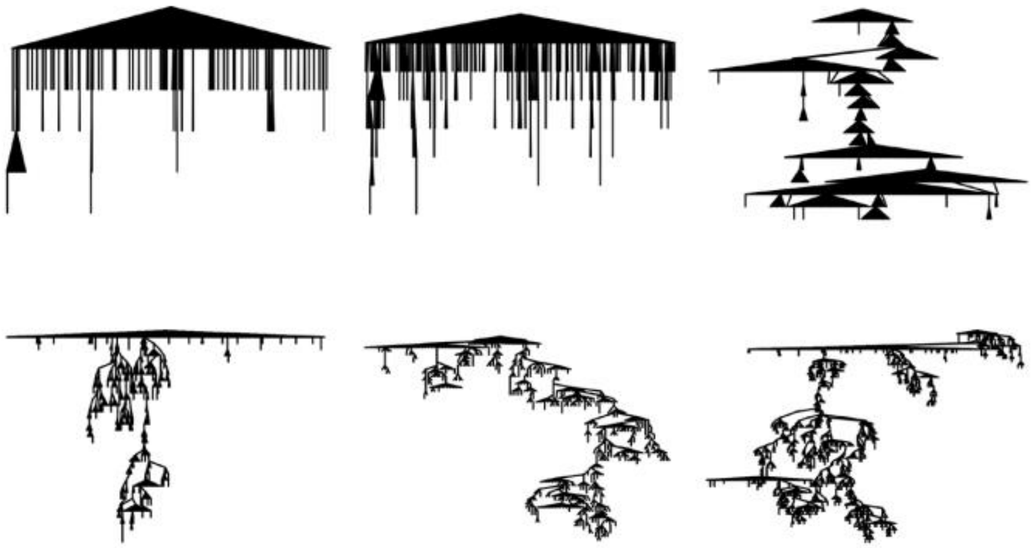
"Figure 2: Three models of information propagation: subscription, network, and algorithm, showing the propagation of one individual post. In the subscription model, the post reaches those who have subscribed to the poster. In the network model, it cascades through the network as long as users who see it choose to further propagate it. In the algorithmic model shown here, users with similar interests (as learned by the algorithm based on their past engagement) are depicted closer to each other. The more similar a user's interests are to the poster's, the more likely they are to be recommended the post. Of course, other algorithmic logics are possible."



Similar to traditional media (broadcast) the user receives content from the creators they are subscribed to. In the 2000s, Facebook and Twitter were subscription models in that users could not reshare or retweet posts. Network models, availing of users' ability to reshare and retweet others' content, create the possibility of virality. Very few platforms are run on a solely algorithmic model. The user sees posts based on predictions of their likelihood to engage with posts. The progression has been from subscription model to network model to algorithm model over the last twenty years. The latter shift has been prominent in Facebook and Instagram's transition, with TikTok's success resulting in pressure for other platforms to keep up. The shift from network to algorithm model creates new challenges for content creators. No longer about building a network of followers, in the algorithmic model, the number of subscribers is irrelevant to how the post will perform. Because success is based on the 'quality' of one individual post, "an algorithm change that devalues a particular type of content could wipe out a creator at any time"

Virality

To measure virality, structural virality trees are used. They are upside down trees that cascade. Structural virality of a post is described as "the number of degrees of separation, on average, between users in the corresponding tree. The deeper the tree, with more branches, the greater the structural virality." (Goel et al)

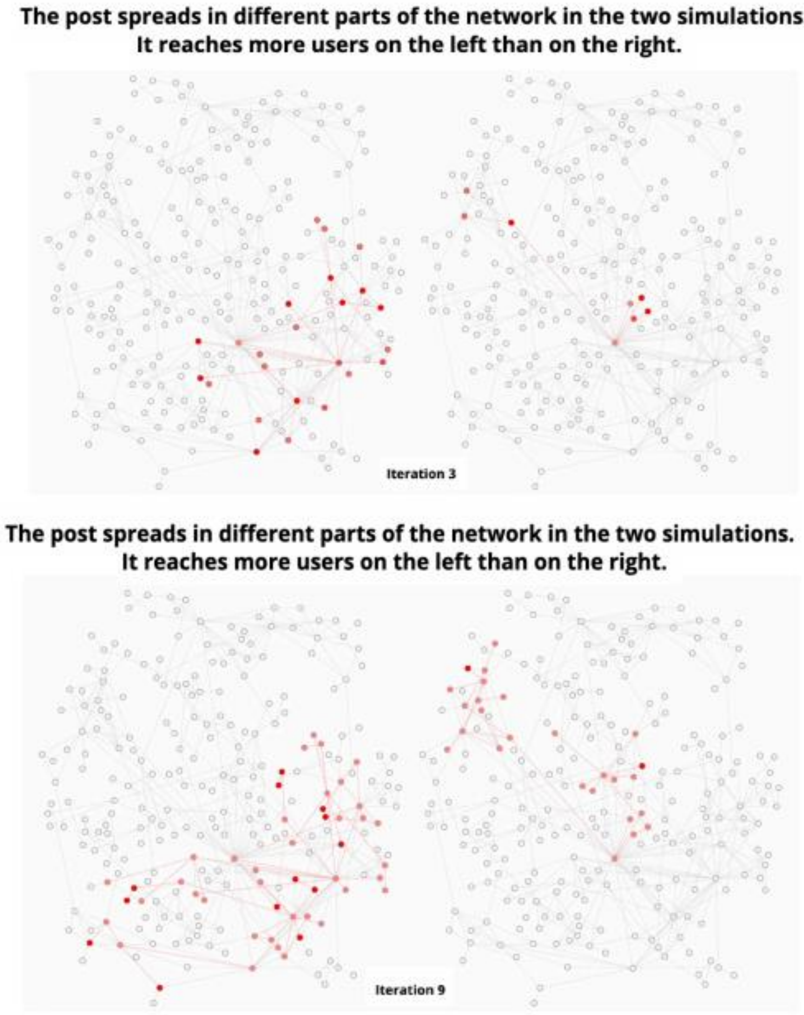


Above are six actual Twitter cascades ordered from least to most viral.

Virality is unpredictable.

"Note that the unpredictability of user behaviour is inevitable. Simply depending on the time of day that a user happens to be on the app, the set of posts they would see in their feed might differ substantially."

Figure 4: Simulation of information cascades in a social network, illustrating the unpredictability of virality.



The post spreads in different parts of the network in the two simulations.
It reaches more users on the left than on the right.

The post spreads in different parts of the network in the two simulations.
It reaches more users on the left than on the right.

Viral content dominates our attention: "less than 1 in 100,000 tweets is retweeted 1,000 times", "The distribution of engagement is highly skewed. A 2022 paper quantified this for TikTok and YouTube: On TikTok, the top 20% of an account's videos get 76% of the views, and an account's most viewed video is on average 64 times more popular than its median video. On YouTube, the top 20% of an account's videos get 73% of the views, and an account's most viewed video is on average 40 times more popular than its median video." (Guinaudeau et al., 2022)

Viral content is highly amenable to demotion

"Demotion, downranking, reduction, or suppression, often colloquially called shadowbanning, is a "soft" content moderation technique in which content deemed problematic is shown to fewer users but not removed from the platform." (Gillespie, 2022). Without demotion, a post would reach most of the network, 10% reduction has little impact on reach, 20% reduction results in a tenfold drop, only reaching the poster's immediate network, outlining the drastic, unpredictable, nonlinear effects demotion has on posts.

As listed by Narayanan: Examples of primary optimization objective for engagement in platforms (that is publicly reported)

- Facebook
    - "'Meaningful Social Interactions', a weighed average of Likes, Reactions, Reshares, and Comments."
- Twitter
    - Similar to Facebook, a combination of all types of interaction a user might have with a tweet.
- YouTube
    - "YouTube optimizes for expected watch time, that is, how long the algorithm predicts the video will be watched."
    - "If a user sees a video in their recommendations and doesn't click on it, the watch time is zero. If they click on it and hit the back button after a minute, the watch time is one minute. Before 2012, YouTube optimized for click-through rate instead, which led to clickbait thumbnails (such as sexualised imagery) becoming ubiquitous; hence the shift to watch time."
- TikTok
    - Less is known about this algorithm than that of other major platforms, but seems to be similarly a combination of liking, commenting, and play time.
    - Videos being watched to completion is a strong signal, and even though the restriction on video length has been lifted, 15 second videos still seem to be most popular (probably because of their relative likelihood to be watched to completion and the subsequent incentivisation by the algorithm).
- Netflix
    - "Netflix originally optimized for suggesting movies that the user is likely to rate highly on a scale of one to five; this was the basis for a $1M recommender system competition, the Netflix Prize, in 2006. But now uses a more complex approach."

Transparency in recommender systems has been implemented by some platforms such as Netflix and Spotify, with explanations of why a recommendation was made being more persuasive. Platform companies seem happy to share information about algorithms in order to learn from others in the industry, but this is from an engineering and research standpoint, with very few sharing information on the impact these algorithms have on their users.

Where does that leave us, the user? A better understanding of these algorithms and their potential effects might alter one's experience online, but in relational privacy fatigue, there are a few measures one can take to limit or avoid algorithmic influence.

1. Completely remove yourself from the internet and clean up your digital footprint.

Cleaning up footprint

Removing old email: Remember each email you have used in the past. You will need to recover them and regain access, to access other websites that may have been signed up to with said email. Searching for name, image search. Delete social media accounts. Removing old accounts and services. Use the search function in email for 'Sign-up', 'welcome' etc. Now browse the service and try to find a way to delete account. Google 'delete account + <service>'. If no luck, email and request to delete account. Visit 'themarkup.org', this website scans websites to show specific user-tracking technologies implemented to harvest user data. webkay.robinlinus.com This website can be used to check what your web browser knows about you

2. Join the 'Dumbphone Revolution'
3. Join the 'Slow Computing Movement'
4. "The Strategy of Non-Participation
   Non-participation doesn't mean abandoning digital platforms entirely. Instead, it's about strategic disengagement: consciously choosing when, where, and how to interact with algorithm-driven systems. Here's how this strategy can empower individuals and businesses:

   1. Reclaim Your Autonomy
   By limiting your interactions with algorithms, you regain control over your choices. For individuals, this might mean curating your own content consumption—seeking out newsletters, direct subscriptions, or independent platforms. For businesses, it could mean prioritizing direct customer relationships through email marketing, in-person events, or exclusive communities.

   2. Focus on Authentic Connections

Algorithms often reward surface-level engagement over meaningful interactions. By stepping away from the algorithm’s demands, you can focus on creating deeper, more authentic connections. Businesses can build loyalty through genuine customer engagement rather than chasing viral moments. Individuals can foster real relationships by spending more time offline or in private, algorithm-free spaces.

3. Diversify Your Presence

Relying solely on algorithmic platforms is risky. A sudden change in the algorithm can tank visibility overnight. Diversify your online presence by exploring decentralized platforms, investing in your own website or blog, or leveraging tools that give you direct control over your audience.

4. Opt for Quality Over Quantity

Algorithms reward frequent posting, often at the expense of quality. By not playing the algorithm’s game, creators can focus on producing thoughtful, high-value content. Businesses can concentrate on fewer but more impactful campaigns. The result? A stronger, more enduring impact.

5. Protect Your Mental Space

Constant participation in algorithm-driven systems can lead to burnout, anxiety, and reduced productivity. Strategic disengagement allows individuals to protect their mental health by reducing information overload and regaining time for reflection and creativity.

Real-World Examples of Non-Participation

Basecamp: The software company famously stepped away from social media marketing to focus on word-of-mouth and direct customer relationships. This move aligned with their values and allowed them to grow sustainably.
Digital Minimalists: A growing community advocates for reducing reliance on algorithmic platforms. They champion intentional tech use, emphasizing tools and practices that serve specific, meaningful purposes.
Independent Creators: Many artists and writers now prioritize newsletters or crowdfunding platforms like Patreon, where they can engage directly with their audience without algorithmic interference.

The Paradox in Practice

> The power of non-participation lies in its paradoxical nature. By opting out, you're no longer feeding the system that demands your attention. This absence can make your presence more valuable, allowing you to stand out in a noisy, oversaturated digital landscape.
>
> Non-participation is not a retreat; it's a strategy. It's about being intentional, prioritizing quality over quantity, and refusing to let algorithms dictate your decisions. In doing so, you can reclaim your time, creativity, and autonomy—and perhaps even inspire others to do the same.
>
> As the digital world grows increasingly algorithmic, the choice to step back becomes not just a personal preference but a powerful act of defiance. After all, to beat the algorithm, the key is not to participate." (David, 2025)

Conclusion

Before starting this project, I had 'folk psychology' notions of how algorithms work. Having explored how some major platforms recommend content, it is evident that the nature of the platform (for example text based or video-streaming) as well as the network of the user (unless the platform has an emphasis on exploration) influence the algorithm's decisions. I hope that this has informed the reader in some way, to better understand what these algorithms strive for, and potential benefits and harms. More research could explore effects of algorithms on people in more countries and across more demographics, including age, gender, ethnicity, education level, occupation, location and possibly a vulnerability metric such as internet access, life experience or presence/lack algorithm awareness. This may prove difficult, with finding a full breadth of participants and the private nature of some internet use.

Bibliography

"Algorithmic Injustice: A Relational Ethics Approach." *Patterns* 2, no. 2 (n.d.). https://doi.org/10.1016/j.patter.2021.100205.

Bakshy, Eytan, Solomon Messing, and Lada A. Adamic. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348, no. 6239 (June 5, 2015): 1130–32. https://doi.org/10.1126/science.aaa1160.

Birhane, Abeba. "Descartes Was Wrong: 'A Person Is a Person through Other Persons.'" *Aeon Magazine*, April 7, 2017. https://aeon.co/ideas/descartes-was-wrong-a-person-is-a-person-through-other-persons.

Bruns, Axel. "Filter Bubble." *Internet Policy Review* 8, no. 4 (November 29, 2019). https://doi.org/10.14763/2019.4.1426.

Buchanan, Tom, Carina Paine, Adam N. Joinson, and Ulf-Dietrich Reips. "Online Privacy and Protection Concerns Measure." *PsycTESTS Dataset*, 2007. https://doi.org/10.1037/t59995-000.

Burgoon, Judee K. "Privacy and Communication." *Annals of the International Communication Association* 6, no. 1 (January 1982): 206–49. https://doi.org/10.1080/23808985.1982.11678499.

Choi, Hanbyul, Jonghwa Park, and Yoonhyuk Jung. "Privacy Fatigue in Online Privacy Behavior Scale." *PsycTESTS Dataset*, 2018. https://doi.org/10.1037/t66187-000.

Dahlgren, Peter M. "A Critical Review of Filter Bubbles and a Comparison with Selective Exposure." Walter de Gruyter, January 29, 2021. https://www.researchgate.net/publication/350174545_A_critical_review_of_filter_bubbles_and_a_comparison_with_selective_exposure.

Darby, Max. "Can Artificial Neural Networks Sufficiently Refute Free-Will?" unknown, April 1, 2018. https://www.researchgate.net/publication/330401471_Can_Artificial_Neural_Networks_Sufficiently_Refute_Free-Will.

David. "The Algorithm's Paradox: Strategic Non-Participation in Digital Platforms." *David's Substack*, January 10, 2025. https://dprlab.substack.com/p/the-algorithms-paradox.

Dienlin, Tobias, and Sabine Trepte. "Is the Privacy Paradox a Relic of the Past? An In-depth Analysis of Privacy Attitudes and Privacy Behaviors." *European Journal of Social Psychology* 45, no. 3 (July 31, 2014): 285–97. https://doi.org/10.1002/ejsp.2049.

Doherty, Brennan. “People Want ‘Dumbphones’. Will Companies Make Them?” *BBC*, May 20, 2024. https://www.bbc.com/future/article/20240515-the-dumbphones-people-want-are-hard-to-find.

Fischer, Peter, Stephen Lea, Andreas Kastenmüller, Tobias Greitemeyer, Julia Fischer, and Dieter Frey. “The Process of Selective Exposure: Why Confirmatory Information Search Weakens over Time.” *Organizational Behavior and Human Decision Processes* 114, no. 1 (January 2011): 37–48. https://doi.org/10.1016/j.obhdp.2010.09.001.

Fouquaert, Thibault, and Peter Mechant. “Making Curation Algorithms Apparent: A Case Study of ‘Instawareness’ as a Means to Heighten Awareness And...” Taylor & Francis, February 28, 2021. https://www.researchgate.net/publication/349698578_Making_curation_algorithms_apparent_a_case_study_of_’Instawareness’_as_a_means_to_heighten_awareness_and_understanding_of_Instagram’s_algorithm.

Frimer, Jeremy A., Harinder Aujla, Matthew Feinberg, Linda J. Skitka, Karl Aquino, Johannes C. Eichstaedt, and Robb Willer. “Incivility Is Rising Among American Politicians on Twitter.” *Social Psychological and Personality Science* 14, no. 2 (April 28, 2022): 259–69. https://doi.org/10.1177/19485506221083811.

Garrett, R. Kelly. “Politically Motivated Reinforcement Seeking: Reframing the Selective Exposure Debate.” *Journal of Communication* 59, no. 4 (December 2009): 676–99. https://doi.org/10.1111/j.1460-2466.2009.01452.x.

Gillespie, Tarleton. “Do Not Recommend? Reduction as a Form of Content Moderation.” *Social Media + Society* 8, no. 3 (July 2022). https://doi.org/10.1177/20563051221117552.

Goel, Sharad, Ashton Anderson, Jake Hofman, and Duncan J. Watts. “The Structural Virality of Online Diffusion.” *Management Science* 62, no. 1 (January 2016): 180–96. https://doi.org/10.1287/mnsc.2015.2158.

Guinaudeau, Benjamin, Kevin Munger, and Fabio Votta. “Fifteen Seconds of Fame: TikTok and the Supply Side of Social Video.” *Computational Communication Research* 4, no. 2 (October 1, 2022): 463–85. https://doi.org/10.5117/ccr2022.2.004.guin.

Kitchin, Rob. “New Book: Slow Computing: Why We Need Balanced Digital Lives.” The Programmable City, n.d. https://progcity.maynoothuniversity.ie/2020/09/new-book-slow-computing-why-we-need-balanced-digital-lives/.

Munson, Sean A., and Paul Resnick. “Presenting Diverse Political Opinions.” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1457–66. New York, NY, USA: ACM, 2010. https://doi.org/10.1145/1753326.1753543.

Narayanan, Arvind. “Understanding Social Media Recommendation Algorithms.” *Knight First Amendment Institute*, n.d. https://knightcolumbia.org/content/understanding-social-media-recommendation-algorithms.

Pariser, Eli. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Press HC, 2011.

Petrovčič, Andraž. “Disentangling the Role of Algorithm Awareness and Knowledge in Digital Inequalities: An Empirical Validation of an Explanatory Model.” *Information, Communication & Society*, March 12, 2025.

Plummer, Libby. “This Is How Netflix’s Top-Secret Recommendation System Works.” *WIRED*, August 22, 2017. https://www.wired.com/story/how-do-netflixs-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like/.

Romanosky, Sasha, Alessandro Acquisti, Jason Hong, Lorrie Faith Cranor, and Batya Friedman. “Privacy Patterns for Online Interactions.” In *Proceedings of the 2006 Conference on Pattern Languages of Programs*, 1–9. New York, NY, USA: ACM, 2006. https://doi.org/10.1145/1415472.1415486.

Siles, Ignacio, and Johan Espinoza Rojas. “How Using Various Platforms Shapes Awareness of Algorithms.” *Behaviour &amp; Information Technology*, January 1, 2022.

Taylor, Dalmas A., and Irwin Altman. “Self-Disclosure as a Function of Reward-Cost Outcomes.” *Sociometry* 38, no. 1 (March 1975): 18. https://doi.org/10.2307/2786231.

Walmsley, J. *Mind and Machine*. Springer, 2016.

Wang, Chenyue, Sophie C Boerman, Anne C Kroon, Judith Möller, and Claes H de Vreese. “The Artificial Intelligence Divide: Who Is the Most Vulnerable?” *New*

Media &amp; Society, February 26, 2024. https://doi.org/10.1177/14614448241232345.