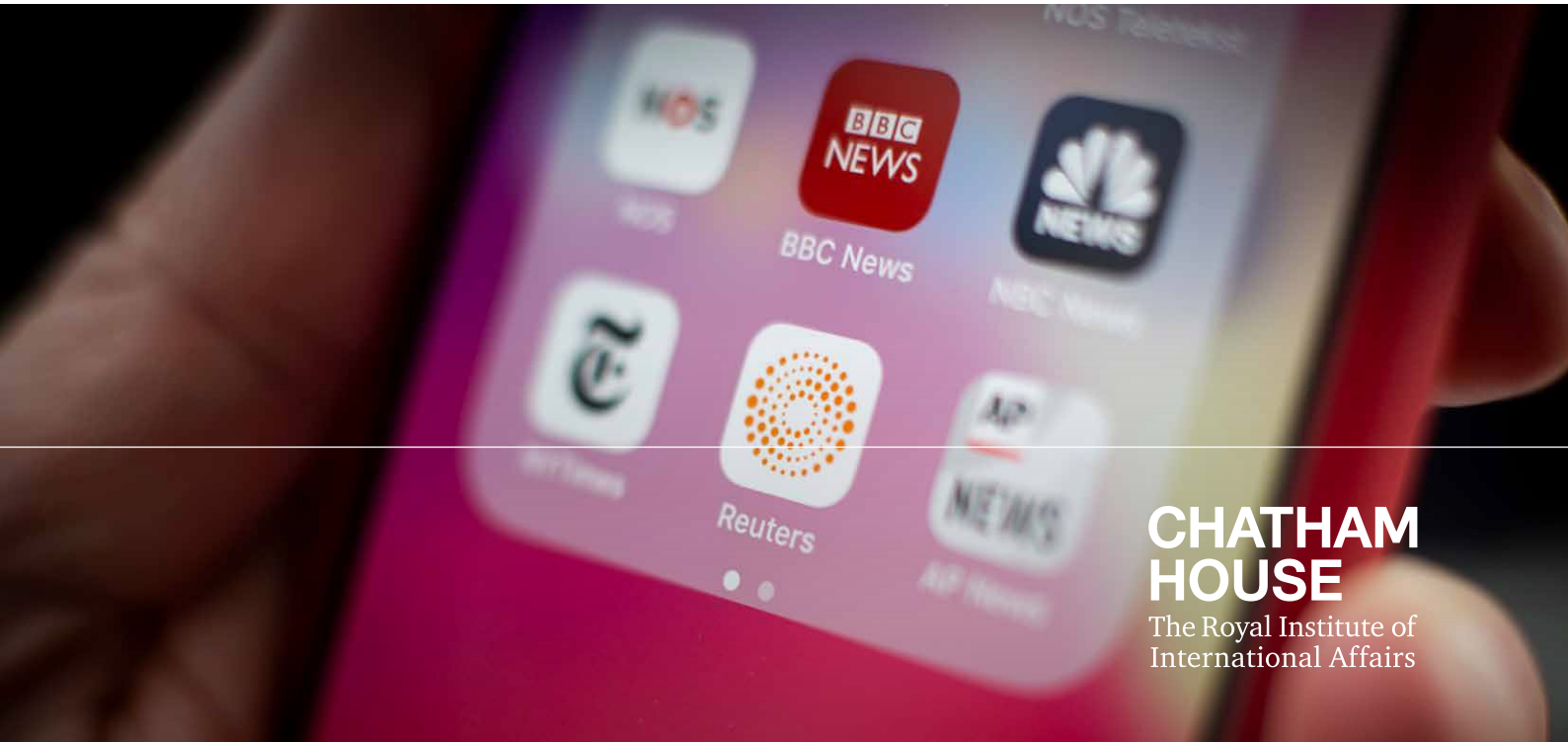
Research Paper

Sophia Ignatidou
International Security Department | December 2019

AI-driven Personalization in Digital Media

Political and Societal Implications



CHATHAM HOUSE
The Royal Institute of International Affairs

Contents

Summary

- Machine learning (ML)-driven personalization is fast expanding from social media to the wider information space, encompassing legacy media, multinational conglomerates and digital-native publishers: however, this is happening within a regulatory and oversight vacuum that needs to be addressed as a matter of urgency.

- Mass-scale adoption of personalization in communication has serious implications for human rights, societal resilience and political security. Data protection, privacy and wrongful discrimination, as well as freedom of opinion and of expression, are some of the areas impacted by this technological transformation.

- Artificial intelligence (AI) and its ML subset are novel technologies that demand novel ways of approaching oversight, monitoring and analysis. Policymakers, regulators, media professionals and engineers need to be able to conceptualize issues in an interdisciplinary way that is appropriate for sociotechnical systems.

- Funding needs to be allocated to research into human–computer interaction in information environments, data infrastructure, technology market trends, and the broader impact of ML systems within the communication sector.

- Although global, high-level ethical frameworks for AI are welcome, they are no substitute for domain- and context-specific codes of ethics. Legacy media and digital-native publishers need to overhaul their editorial codes to make them fit for purpose in a digital ecosystem transformed by ML. Journalistic principles need to be reformulated and refined in the current informational context in order to efficiently inform the ML models built for personalized communication.

- Codes of ethics will not by themselves be enough, so current regulatory and legislative frameworks as they relate to media need to be reassessed. Media regulators need to develop their in-house capacity for thorough research and monitoring into ML systems, and – when appropriate –proportionate sanctions for actors found to be employing such systems towards malign ends. Collaboration with data protection authorities, competition authorities and national electoral commissions is paramount for preserving the integrity of elections and of a political discourse grounded on democratic principles.

- Upskilling senior managers and editorial teams is fundamental if media professionals are to be able to engage meaningfully and effectively with data scientists and AI engineers.

1. Introduction

The so-called '50/50' moment, when more than half of the world's population was online, was reached in 2018,[1] but the current digital ecosystem struggles to reconcile the geopolitics of information with human rights protections; the sovereignty of government with the market imperatives of technology companies; and, perhaps more insidiously, normative shifts with the stability of democracies and belief in truth itself. The fallout from disinformation and online manipulation strategies have alerted Western democracies to the novel, nuanced vulnerabilities of our information society. And influence operations threatening the political security[2] of states around the world have brought the concept of 'information wars' into sharp focus. Opinion surveys indicate that citizens expect journalists to tackle disinformation,[3] but this information crisis has found the Fourth Estate (the global press and news media) exhausted financially – with the majority of digital advertising spending currently being channeled towards the 'big three' technology companies of Google, Facebook and Amazon[4] – and with its status as a pillar of democracy undermined by attacks that have come to be known in popular discourse as 'fake news'.

What happens in journalism matters, because its public-service role of holding power to account is fundamental for democracy. But for that role to remain tenable, journalism has to ride the latest wave of technological transition on its own terms, and in ways that align with its core purpose, values and priorities, which will need to be refined and re-examined in the digital context. For more than a decade technology companies – unrestrained and unregulated – have been unilaterally restructuring the digital space, in ways that were based on unchallenged sets of assumptions and that made legacy media[5] dependent on really powerful forces outside the newsroom.[6] This kind of 'infrastructural capture'[7] raises concerns in relation to the latest technology to take over newsrooms: artificial intelligence (AI), and its machine learning (ML) subset.

Ethics are part of a broader effort to resolve the complex tensions that continue to arise, but they should be seen as just one of the parameters of the adaptive, human-centric, sustainable, accountable and resilient framework that needs to be set in place. AI, together with the various levels and variants of personalization it enables, necessitates a delicate balancing between ethical, political and societal concerns on the one hand, and consumers' and markets' needs on the other. AI is a rapidly evolving field, shaped by various and often contesting political and economic normative powers. The speed of its development demands that policymakers embrace forward-looking

[1] International Telecommunications Union (2018), 'ITU releases 2018 global and regional ICT estimates', 7 December 2018, https://www.itu.int/en/mediacentre/Pages/2018-PR40.aspx (accessed 11 Sept. 2019).
[2] This paper adopts Barry Buzan's definition of political security, as a term that 'concerns the organizational stability of states, systems of government, and the ideologies that give them legitimacy'. See Buzan, B. (1991), 'New patterns of global security in the twenty-first century', *International Affairs*, 67(3): p. 433.
[3] European Commission, Brussels (2018), *Flash Eurobarometer 464:* (*Fake News and Disinformation Online*, TNS opinion, Brussels [producer], GESIS Data Archive, Cologne, ZA6934 Data file Version 1.0.0, doi: 10.4232/1.13019, April 2018, p. 4.
[4] Oreskovic, A. (2019), 'This chart shows just how much Facebook, Google, and Amazon dominate the digital economy', Business Insider, 16 June 2019, https://www.businessinsider.com/facebook-google-amazon-dominate-digital-economy-chart-2019-6 (accessed 1 Jul. 2019).
[5] Legacy media denotes mass media, such as traditional broadcasters and newspapers. In the UK, the *Guardian* and the BBC are obvious examples.
[6] For instance, news publishers' posts on Facebook's News Feed were deprioritized in 2018, and the company's prioritization of native videos in 2014 also had an immediate impact on the output of news. For the latter case, see Maitra, J. and Tandoc, Jr, E. C. (2018), 'News organizations' use of Native Videos on Facebook: tweaking the journalistic field one algorithm change at a time', *New Media & Society*, 20(5): pp. 1679–96.
[7] Nechushtai, E. (2017), 'Could digital platforms capture the media through infrastructure?', *Journalism*, 19(8).

approaches such as foresight analysis, and make impact assessments a core component of their toolkit.

This paper seeks to outline the implications of the adoption of AI, and more specifically of ML, by the old 'gatekeepers' – the legacy media – as well as by the new, algorithmic, media – the digital intermediaries – focusing on personalization. Data-driven personalization, despite demonstrating commercial benefits for the companies that deploy it, as well as a purported convenience for consumers, can have individual and societal implications that convenience simply cannot counterbalance. Nor are citizens necessarily complacent with regard to targeting, as has been suggested. According to an interim report on online targeting released by the UK's Centre for Data Ethics and Innovation (CDEI), 'people's attitudes towards targeting change when they understand more of how it works and how pervasive it is'.[8]

Methodology

This paper is the outcome of a wide-ranging literature review covering academic papers, policy documents by research institutions and international organizations, panel discussions at the CPDP (Computers, Privacy and Data Protection) 2019 conference in Brussels, the Data Science Salon series in the US, and the CogX 2019 conference in London. It has also profited immensely from interviews with professionals from journalism, data science and technology companies, as well as policymakers and researchers. Space and language limitations have meant that the list of companies and media represented here is far from exhaustive, but the paper has the broader aim of identifying the long-term implications of personalization in digital communications for political security, citizens' autonomy, journalism and public discourse, and to contribute to a more extensive process of research by the appropriate stakeholders. In terms of personalization in legacy and social media, the analysis relates to the dissemination of breaking news and political content (termed 'hard' news), rather than 'soft' news such as entertainment. Personalization is employed by a growing number of companies, so even though we are focusing on technology and media companies with dominant or substantial market power, the observations made by this research relate to communication actors not covered by this definition. The landscape is moving extremely rapidly, so it is important to note that this paper relates to the prevailing situation in late November 2019.

[8] UK Department for Digital, Culture, Media and Sport, Centre for Data Ethics and Innovation (2019), *Interim report: Review into online targeting*, 25 July 2019, p. 7, https://www.gov.uk/government/publications/interim-reports-from-the-centre-for-data-ethics-and-innovation/interim-report-review-into-online-targeting (accessed 11 Sept. 2019).

2. Terminologies and Taxonomies

ML-driven personalization is the latest application of a technology that has started transforming a number of services and industries, from health to retail, hospitality and journalism. The two prevailing forms of personalization in digital communication are targeting and recommender systems, although rapid advances in natural language processing (NLP), synthetic media (so-called 'deepfakes', for example), conversational journalism, and 'Internet of Things' (IoT) devices are expected to enable more granular, multi-platform and interactive customization. Conversational AI systems such as IBM's Debater[9] will be able to engage into debates with individuals in a targeted manner. Even though a recent study by Google AI on a universal neural machine translation system concluded that an effective system translating between hundreds of languages is still some way off,[10] it may become commonplace in the future. Given the scope of the move from mass media communication to personalization, there is great value in examining closely the state of play in the latter, by first defining what it entails:

- **Personalization** is the customization of content to the individual through engagement in information filtering, classifying, prioritizing and adjusting. It can be explicit, using direct user inputs, or implicit, drawing on inferences created by the data.[11] For others, the difference between implicit and explicit personalization relates to self-selected or default personalization.[12] This paper adopts the latter definition, although the former will demand our attention as use of inferences becomes more widespread.

- **Targeting** is a form of personalization. On the basis of profiling, individuals are targeted with personalized content that is expected to have a specific impact on their decisions or behaviour. Profiling is enabled by the tracking of digital trails and bulk data collection, through the use of cookies, social plug-ins, tracking pixels, ambient sensors (using Wi-Fi or Bluetooth),[13] or third-party code embedded in applications. The current data-mining explosion has empowered targeted advertising that takes in geolocation, IP addresses, browsing histories, and information mined from IoT devices.[14] Targeting can take many forms – direct, location-based, contextual or demographic.[15]

[9] Metz, C. and Lohr, S. (2018), 'IBM Unveils System That "Debates" With Humans', *New York Times*, 18 June 2018, https://www.nytimes.com/2018/06/18/technology/ibm-debater-artificial-intelligence.html (accessed 25 Jun. 2019).
[10] Arivazhagan, N., Bapna, A., Firat, O., Lepikhin, D., Johnson, M., Krikun, M., Chen, M. X., Cao, Y., Foster, G., Cherry, C., A. Macherey, W., Chen, Z. and Wu, Y. (2019), '*Massively Multilingual Neural Machine Translation in the Wild: Findings and Challenges*', 11 July 2019, https://arxiv.org/abs/1907.05019 (accessed 11 Sept. 2019).
[11] Thurman, N. and Schifferes, S. (2012), 'The future of personalization at news websites: Lessons from a longitudinal study', *Journalism Studies*, 13(5–6): pp. 775–90, doi: 10.1080/1461670X.2012.664341 (accessed 11 Sept. 2019).
[12] Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., de Vreese, C. H. and Helberger, N. (2016), 'Should we worry about filter bubbles?', *Internet Policy Review*, 5(1), doi: 10.14763/2016.1.401 (accessed 11 Sept. 2019).
[13] Mavroudis, V. and Veale, M. (2018), 'Eavesdropping whilst you're shopping: Balancing personalisation and privacy in connected retail spaces', Proceedings of the PETRAS/IoTUK/IET, Living in the Internet of Things Conference, London, 28–29 March 2018.
[14] The Internet of Things (IoT) broadly refers to sensors and devices that communicate and collect data. A thorough examination of the capabilities of targeting can be found in Bartlett, J., Smith, J. and Acton, R. (2018), *The Future of Political Campaigning*, London: Demos, July 2018, https://demosuk.wpengine.com/wp-content/uploads/2018/07/The-Future-of-Political-Campaigning.pdf (accessed 1 Jul. 2019). Also for the UK context see ICO's *Investigation into the Use of Data Analytics in Political Campaigns: a Report to Parliament*, 6 November 2018, https://ico.org.uk/media/action-weve-taken/2260271/investigation-into-the-use-of-data-analytics-in-political-campaigns-final-20181105.pdf (accessed 1 Jul. 2019).
[15] Cobbe, J. (2019), 'Panel: Law and policies around targeting', Workshop on the Methodology and Ethics of Targeting, closed workshop presentation, 10 May 2019, Leverhulme Centre for the Future of Intelligence.

- **Recommending** works on the basis of filtering, ranking and prioritizing of content. Filtering can operate on the basis of *popularity*, or it can be *semantic* (based on users' previous online behaviour) and *collaborative* (based on the preferences of segmented audiences to which users belong).[16] In general, popularity and novelty[17] tend to play an important role in recommender systems. Recommending is distinct from targeting, as Cobbe and Singh explain: 'the active and deliberate selection of particular audiences or categories of audience by advertisers is the key distinguishing point between recommending and targeting'.[18] This paper also adopts Cobbe and Singh's taxonomy of *open recommending* (user-generated content distribution such as Google Search or Facebook's News Feed), *curated recommending* (such as Netflix's library), and *closed recommending* (when the content distributed is created by the same organization, such as in the case of the *New York Times* recommender system).

AI, ML and algorithmic gatekeepers

> Technology is neither good nor bad; nor is it neutral.
> – Melvin Kranzberg, 1986[19]

Although there is no widely accepted definition of AI, the UK Government Office for Science's definition is useful, defining AI as the 'analysis of data to model some aspect of the world. Inferences from these models are then used to predict and anticipate possible future events.'[20] AI is broadly divided into two categories, narrow and general, with ML being a subset of the former. ML can create adaptive systems able to improve over time by recognizing patterns in datasets without being explicitly programmed.[21] Algorithms are core components of computational processes and AI applications such as ML, as they comprise 'a sequence of instructions that are carried out to transform the input to the output'.[22]

AI development has been boosted by deep learning – the processing of vast quantities of data via non-linear neural networks that classify the outputs at different layers, creating a complex structure that ultimately functions as a 'black box'.[23] The complexity of black box algorithms has implications for the ability of organizations to be transparent about their handling of data, and researchers have

[16] Möller, J., Trilling, D., Helberger, N. and van Es, B. (2018), 'Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity', *Information, Communication & Society*, 21 (7): pp. 959–77, doi: 10.1080/1369118X.2018.1444076 (accessed 11 Sept. 2019).
[17] Prioritizing popularity may leave out complicated but important news, while novelty certainly did not favour the Occupy Wall Street protesters, for example, as Tarleton Gillespie indicated. See Gillespie, T. (2012), 'Can an algorithm be wrong?', Limn (2), https://limn.it/articles/can-an-algorithm-be-wrong.
[18] Cobbe, J. and Singh, J. (2019), 'Regulating Recommending: Motivations, Considerations, and Principles', 15 April 2019, SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3371830 (accessed 3 Jul. 2019), p. 8.
[19] Kranzberg, M. (1986) 'Technology and History: "Kranzberg's Laws"', *Technology and Culture*, 27(3): pp. 544–60.
[20] Government Office for Science (2016), *Artificial intelligence: opportunities and implications for the future of decision making*, 12 February 2016, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf (accessed 1 Jul. 2019). Also useful is the definition used by the UK government in its *Industrial Strategy: Building a Britain fit for the future* (2017), defining AI as 'technologies with the ability to perform tasks that would otherwise require human intelligence, such as visual perception, speech recognition, and language translation'.
[21] This learning can be supervised, using labelled training datasets that include both the input and the target output for each instance, unsupervised, left to spot regularities and patterns in the data without providing any feedback to the system; or, in the case of reinforcement learning, systems of rewards can drive algorithms towards a certain goal. Google's original PageRank algorithm is an example of unsupervised learning.
[22] Alpaydin, E. (2016), *Machine Learning*, The MIT Press Essential Knowledge Series, Cambridge, MA: MIT Press, p. 16.
[23] Pasquale, F. (2015), *The Black Box Society: The Secret Algorithms that Control Money and Information*, Cambridge, MA: Harvard University Press.

warned that what big data[24] often offers is 'the power to predict without understanding'.[25] Adding to the confusion is the idea of big data being coterminous with AI. Exponential amounts of data do not necessarily mean that there are underlying rules that can be gleaned. Algorithmic models are built on dependency assumptions that have to be accurate, otherwise pattern correlations might be misconstrued as causation. Not all models resemble a 'black box' – the selection of the model (linear regression, random forests, etc.) will impact its explainability.

ML algorithmic systems differ in terms of models, the performance criteria they optimize for, and the way each model's parameters are adjusted.[26] In digital communication, online engagement – measured by the click-through rate (CTR) – is the prevalent metric for which most social and legacy media optimize, although the parameters and models they employ may vary.

In the information space, ML algorithms are gradually playing the role of gatekeeper by filtering out information, essentially discriminating against certain content. It is this role that some argue should constitute digital platforms operating recommendation systems exempt from liability protections, such as those provided under the EU's E-Commerce Directive.[27] In the US, digital intermediaries are protected by Section 230 of the Communications Decency Act, although the opening of antitrust investigations in mid-2019 at the House of Representatives[28] and, more recently, by a series of US states,[29] indicates that US policymakers may be ready to review that legal immunity. In late November, moreover, Democrats in the Senate proposed wide-ranging federal data privacy legislation – termed the Consumer Online Privacy Rights Act – intended to 'provide consumers with foundational data privacy rights, create strong oversight mechanisms, and establish meaningful enforcement'.[30]

The gatekeeping role of current digital intermediaries cannot be overstated. Across all countries surveyed by the 2019 Reuters Institute Digital News Report, just 29 per cent of interviewees said they preferred to access a news website directly, with 55 per cent of the combined sample stating they preferred to access news through search engines, social media and news aggregators,[31] and with younger users being more likely to use social media and aggregators. The report also noted that mobile aggregation is the 'majority behaviour' in many Asian countries.[32] Aggregators' market share is another aspect meriting attention. In the US, for example, Apple News reaches more iPhone users (27 per cent) than the Washington Post (23 per cent).[33] But, in a world of algorithmic

[24] A usual misconception is that big data is fundamental for deep learning. Although vast datasets can be used, they are not necessary.
[25] Gorton, W. A. (2016), 'Manipulating Citizens: How Political Campaigns' Use of Behavioral Social Science harms Democracy', *New Political Science*, 38(1): p. 73.
[26] Alpaydin (2016), *Machine Learning*, p. 39.
[27] Cobbe and Singh (2019), 'Regulating Recommending: Motivations, Considerations, and Principles'.
[28] Isaac, M., Lohr, S. and Popper, N. (2019), 'Tech Hearings: Congress Unites to Take Aim at Amazon, Apple, Facebook and Google', *New York Times*, 16 July 2019, https://www.nytimes.com/2019/07/16/technology/big-tech-antitrust-hearing.html (accessed 11 Sept. 2019).
[29] Paul, K. (2019), 'Facebook and Google antitrust investigations: all you need to know', *Guardian*, 7 September 2019, https://www.theguardian.com/technology/2019/sep/06/facebook-google-antitrust-investigations-explained (accessed 11 Sept. 2019).
[30] Rushe, D. (2019, 'Democrats propose sweeping new online privacy laws to rein in tech giants', *Guardian*, 26 November 2019, https://www.theguardian.com/world/2019/nov/26/democrats-propose-online-privacy-laws (accessed 2 Dec. 2019). For the full text of the draft bill, see https://www.cantwell.senate.gov/imo/media/doc/COPRA%20Bill%20Text.pdf.
[31] Newman, N., Fletcher, R., Kalogeropoulos, A. and Nielsen, R. K. (2019), *Reuters Institute Digital News Report 2019*, Reuters Institute for the Study of Journalism, Oxford: University of Oxford, p. 13, https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-06/DNR_2019_FINAL_0.pdf (accessed 30 Jun. 2019).
[32] Ibid., p. 17.
[33] Ibid., p. 10.

gatekeeping, a key question raised by some observers is: what kind of gatekeepers do we want algorithmic systems to be?[34]

[34] Nechushtai, E. and Lewis, S. C. (2019), 'What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations', *Computers in Human Behavior*, 90: pp. 298–307.

3. Big Tech, Big Data, Big Issues

> As this media segmentation advances, people will not develop any common fund of knowledge. They will become unable to engage in civic talk; they will have nothing to say to each other. Any common public sphere will wither and die.
> – C. Edwin Baker, 1998.[35]

Large technology companies started using personalization as a way to manage the exponential increase of data collected by their systems. But social media's drive to harness communication was heavily dependent on their imperative to monetize it.[36] A 2010 study showed that, while advertising rates for behaviourally targeted advertisements[37] may be higher than 'run-of-network' (RON) advertising,[38] the former's conversion rates are over twice as high as those of the latter.[39] The ad auctions operated by social media companies[40] profit immensely from personalization as the auctions enable sites to charge more[41] in a marketplace where bidders compete for attention. This 'attention economy' infrastructure nudges users to share an increasing amount of personal data that in turn feeds back into the system, by appealing to their neural reward systems.[42] The irony of putting the burden of 'informed consent' on users who are asked to allocate a disproportionate amount of their time to reading terms and conditions, while at the same time extracting as much attention from them as possible, should not elude us.

Personalization is more than a marketing strategy. It can become an important enabler in evaluating, predicting and potentially reorienting the behaviour of large user groups, not just in terms of their relationship with advertisers but with the tech companies themselves, who can use this information to enhance their market position. For revenue streams to continue to scale, it is imperative to maintain not only a client base of advertisers but also a committed user base whose behaviour can be modelled. In relation to social media's filtering algorithmic systems, Stuart Russell, a leading AI expert, has suggested that, given the goal of maximizing CTR, systems' priority might not be to serve content that users are likely to engage with, but to shape their preferences so

[35] Baker, C. E., (1998), 'Media That Citizens Need', *University of Pennsylvania Law Review*, 147(2): p. 365.
[36] Social media often collect data not just in 'active' but also in 'passive' ways, even from non-users. For instance, for a look at Google's data collection practices see Schmidt, D. C. (2018), 'Google Data Collection', Digital Content Next, 21 August 2018, https://digitalcontentnext.org/blog/2018/08/21/google-data-collection-research (accessed 8 Jun. 2019).
[37] Behavioural targeting uses a series of signals, mainly from users' previous online activity – clicks, websites visited, and more – to estimate, through the use of models, which ads those users will be more responsive to.
[38] Run of network advertising applies an online campaign randomly to various sites across an ad network, as opposed to the predetermined placement of targeted ads.
[39] Beales, H. (2010), 'The value of Behavioral Targeting', https://www.networkadvertising.org/pdfs/Beales_NAI_Study.pdf (accessed 11 Sept. 2019).
[40] For Facebook, see https://www.facebook.com/business/help/163066663757985; for Google, https://support.google.com/adsense/answer/160525?hl=en-GB; and for Amazon, https://sellercentral.amazon.com/gp/help/external/G201528470?language=en_US.
[41] Ad auctions operate on the basis that service provides (for example social media companies), enable different bidders to compete for digital ads predicted to lead to interactions, based on their targeting to profiled groups or individuals. The logic is that enhanced profiling and refined targeting leads to more engagement, allowing ad providers to charge more.
[42] In April 2019 US Senators Mark Warner and Deb Fischer introduced bipartisan legislation, the Deceptive Experiences to Online Users Reduction (DETOUR) Act, to ban the use by social media platforms of 'dark patterns' which aimed to coerce users to share more of their data or expose themselves to exploitation, although some of the Act's clauses – such as banning audience segmentation without consent – are ambitious in the current context. The Subcommittee on Communication, Technology, and Innovation has also started looking into persuasive technologies. https://www.warner.senate.gov/public/index.cfm/2019/4/senators-introduce-bipartisan-legislation-to-ban-manipulative-dark-patterns (accessed 24 Jul. 2019).

their behaviour can become more predictable.[43] Looking ahead, we have to acknowledge that AI systems may be solving problems in their own unique and at times unpredictable ways.

Digital intermediaries' underlying business models are crucial because they dictate their architecture – which in turn regulates the norms they propagate. These corporate entities can control and reshape the infrastructure of public discourse, and by extension the environment affecting democratic elections.[44] They shape the way campaigns perceive the electorate,[45] and they set the rules based on which political actors appeal to voters[46] – A/B ad testing,[47] for example, could lead to prioritizing digital microtargeting rather than more collective real-world campaign events such as rallies, while the virality of emotional content boosted by algorithmic systems can lead to more emotionally charged political discourse too.

Contrary to the claims of technological utopianism, technology that enhances communication is not inherently democratic, and the vulnerabilities of digital intermediaries' architecture to manipulation have been shown on various occasions.[48,49,50] Research has suggested that algorithmic personalization methods, in particular, can enable the manipulation of individual and group opinion dynamics.[51]

The lack of transparency about how tech companies' algorithmic systems operate may reinforce the assumption that their output is an objective representation of reality. Research has indicated that so-called 'black hat'[52] search engine optimization techniques or other manipulation effects (SEME) are possible[53] and can alter voting behaviour. It has been suggested that 'information gerrymandering' can sway vote outcomes by the use of strategically positioned 'zealot' agents targeting voters who have been identified as persuadable.[54] Publicly accessible, adaptable collaborative recommender systems can also be vulnerable to adversarial attacks attempting to co-opt them.[55]

[43] Russell, S. (2019), *Human Compatible: AI and the Problem of Control*, Allen Lane, 2019, p. 8.
[44] Nemitz, P. (2018), 'Constitutional democracy and technology in the age of artificial intelligence', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, http://dx.doi.org/10.1098/rsta.2018.0089 (accessed 8 Jun. 2019).
[45] Data-informed digital campaigning in general has enhanced the perception voter behaviour can be modelled and predicted, leading parties to focus particularly on marginal or 'swing' seats for example as opposed to the public at large.
[46] Kreiss, D. (2019), 'Panel: Data is (political) power!', CPDP 2019 Conference, Brussels, 30 January 2019, https://www.youtube.com/watch?v=TopOHzLx3CE (accessed 8 Jul. 2019).
[47] 'Split', A/B or single variate testing in digital marketing tests two variations of content and its impact on audience, while multivariate testing includes more layered audience segmentation and encompasses more variables, providing more insights into user behaviour.
[48] Nadler, A., Crain, M. and Donovan, J. (2018), *Weaponizing the Digital Influence Machine: the Political Perils of Online Ad Tech*, Data & Society, 17 October 2018, https://datasociety.net/output/weaponizing-the-digital-influence-machine (accessed 8 Jun. 2019).
[49] For instance, two researchers were able to showcase that Facebook ad targeting could be refined to reach a specific individual, although after being notified, Facebook fixed the vulnerability. Faizullabhoy, I. and Korolova, A. (2018), 'Facebook's Advertising Platform: New Attack Vectors and the Need for Interventions', Workshop on Technology and Consumer Protection, 24 May 2018, San Francisco, https://arxiv.org/abs/1803.10099 (accessed 4 Jul. 2019).
[50] The Redirect Method, initially used by Google to redirect to ads aiming to deradicalize extremists, could be utilized by malicious actors, as the New York Times reported. See Berlinquette, P. (2019), 'I Used Google Ads for Social Engineering. It Worked.', *New York Times*, 7 July 2019, https://www.nytimes.com/2019/07/07/opinion/google-ads.html (accessed 8 Jul. 2019).
[51] Perra, N. and Rocha, L. E. C. (2019), 'Modelling opinion dynamics in the age of algorithmic personalisation', *Nature: Scientific Reports 9*, https://www.nature.com/articles/s41598-019-43830-2#Bib1 (accessed 4 Jul. 2019).
[52] In general in computer science, 'black hat' operations indicate aggressive actors and strategies aiming to take advantage of a system's vulnerabilities for malicious aims. In contrast, 'white-hat' researchers tend to probe a system, to expose vulnerabilities with the ultimate goal of finding solutions and enhancing security.
[53] Epstein and Robertson conducted experiments in the US that indicated that biased rankings could affect the outcome of elections. See Epstein, R. and Robertson, R. E. (2015), 'The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections', *Proceedings of the National Academy of Sciences* 112(33): https://www.pnas.org/content/112/33/E4512 (accessed 11 Sept. 2019).
[54] Stewart, A. J., Mosleh, M., Diakonova, M., Arechar, A. A., Rand, D. G. and Plotkin, J.B. (2019), 'Information gerrymandering and undemocratic decisions', *Nature* 573: pp. 117–21, https://www.nature.com/articles/s41586-019-1507-6 (accessed 11 Sept. 2019).
[55] Mobasher, B., Burke, R., Bhaumik, R. and Williams, C. (2007), 'Towards Trustworthy Recommender Systems: An Analysis of Attack Models and Algorithm Robustness', *ACM Transactions on Internet Technology (TOIT),* 7(4):23, pp. 1–41.

Policymakers need to systematically investigate the infrastructure underpinning so-called 'Web 2.0' sites and applications (i.e. based on interactivity and user-generated content), as it is built on unsolicited sets of assumptions about importance – reflected in the weighting of the recommending parameters, for example – that are highly domain- and context-specific to the company that produces the algorithmic systems.[56] Social media's ability to monetize CTR as a quantifiable proof of engagement and harness the affiliated data to improve their ad delivery systems set in motion unprecedented network effects, securing increasing ad revenue and market power. Soon, clicks became the de facto value in the digital space, as big tech firms, seen as the prodigies of the internet era, managed to grow exponentially on the basis of that metric. Still, the legitimacy of the creation of social graphs owned and monitored by private actors such as tech companies, that inform recommendations and targeted advertising is yet to be fully explored.

While social media have solved collective action problems by enabling coordination and have assisted legacy media to scale up their reach by creating new digital dissemination channels, the trade-offs inherent in the way they restructure communication have not been properly examined. Values-in-design – the movement highlighting how technology design can embed normative values – can present practical challenges, but as a principle it merits attention. Taina Bucher, in a 2011 study of Facebook's EdgeRank algorithm, argued that algorithms 'occupy a peculiar epistemological position', by rendering certain elements visible while obscuring others.[57] Dominant technology companies' unprecedented scale, power and unbridled expansion[58] call for reflection on the risks they pose and an appraisal of mitigation options, while their dominant market status continues to propel them at the forefront of AI development.

Open recommending

Google's mission statement, 'to organize the world's information and make it universally accessible and useful',[59] on the face of it is a noble one, but ambitions of global scale tend to demand closer scrutiny. Google Search, owned by Alphabet Inc., is an open recommender system that uses various signals to personalize searches, such as location, previous search keywords and more, from which users can opt out.[60] Similarly, the news aggregator Google News draws on users' activity across other Google applications, such as email, browsing, location services, calendars and more. Since September 2018 the company has also offered personalization solutions to other companies via its Google Optimize service.[61]

[56] For instance, the fact that the nations currently leading in AI development are the US and China will affect the way the prevailing algorithmic systems operate.
[57] Bucher, T. (2012), 'Want to be on the top? Algorithmic power and the threat of invisibility on Facebook', *New Media & Society*, 14 (7): p. 1172.
[58] For instance, Alphabet Inc., Google's parent company, has launched a 'smart city' subsidiary, Sidewalk Labs; Amazon is participating in a healthcare joint venture company, Haven, with Berkshire Hathaway Inc and JPMorgan Chase & Co, and Facebook has announced the proposed launch of its own cryptocurrency, Libra, in 2020.
[59] Google Inc. (2019), 'About', https://about.google/intl/en_us.
[60] Users can change their YouTube, location, web and app activity controls here: https://myaccount.google.com/activitycontrols.
[61] Iziduh, R. (2018), 'Personalization features now available in Google Optimize', Google Blog, 25 September 2018, https://www.blog.google/products/marketingplatform/analytics/personalization-features-now-available-google-optimize (accessed 30 Jun. 2019).

Similarly, in mid-2019 Facebook released its open-source Deep Learning Recommendation Model[62] for personalized content delivery. These personalization offerings – and their successors – will need to be monitored, as, when widely used, privately created AI packages can come to perform a 'standardizing role'.[63] Facebook's business model relies on the constant, cross-device and cross-platform recording of social gestures that create the social graph which can then add a premium to its advertising sales. 'Likes', 'comments' and 'shares' become digital objects, logging implicit or explicit information about each user's age, gender, location, interests and behaviour. These, in turn, are used to refine the user's personalized News Feed, as well as Facebook's targeted ads. Facebook's News service will also include personalized curated recommendations for readers.[64] Twitter's timeline is also a personalized recommender system, from which users can opt out.[65]

Targeted advertising

Digital advertising expenditure is expected to surpass traditional adspend in a number of countries in 2019.[66] Indicatively, candidates in the 2016 US electoral cycle spent almost 800 per cent more on digital advertising than the candidates did in 2012,[67] while in the campaign for the general election held in the UK in 2017, digital ad expenditure reached 42.8 per cent of the total.[68] In particular, microtargeting[69] has been used by political parties of various persuasions in the UK.[70] In the days immediately after Boris Johnson took office as prime minister, for instance, the Conservatives were reported to have posted multiple versions of his commitment to 'deliver Brexit' by 31 October in order to test the efficacy of each version with recipients.[71] And in the lead-up to the UK general election due to take place in December 2019, targeted political advertising was being used by various parties,[72] with legacy media outlets including the UK's Channel 4[73] and the *Guardian*[74] urging the public to supply information on political advertising that they consider to be manipulating or not transparent. Granular analysis of user data (via A/B or 'split' testing, for example) enables a more refined and gradually automated optimization of targeting. Aspects that

[62] Johnson, K. (2019), 'Facebook open-sources DLRM, a deep learning recommendation model', Venture Beat, 2 July 2019, https://venturebeat.com/2019/07/02/facebook-open-sources-dlrm-a-deep-learning-recommendation-model (accessed 3 Jul. 2019).
[63] Cihon, P. (2019), *Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development*, Future of Humanity Institute, University of Oxford, April 2019, p. 24, https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-FHI-Technical-Report.pdf.
[64] Kafka, P. (2019), 'Facebook has finally decided that the best way to deliver news is to act like a newspaper', *Vox*, 25 October 2019, https://www.vox.com/recode/2019/10/25/20931407/facebook-news-tab-newspapers-journalism (accessed 17 Nov. 2019).
[65] Users can opt out here: https://twitter.com/personalization.
[66] Enberg, J. (2019), 'Global Digital Ad Spending 2019', eMarketer, 28 March 2019, https://www.emarketer.com/content/global-digital-ad-spending-2019 (accessed 28 Jun. 2019).
[67] Ghosh, D. and Scott, B. (2018), *Digital Deceit II: A Policy Agenda To Fight Disinformation on the Internet*, New America & The Shorenstein Center on Media, Politics and Public Policy, September 2018, p. 9. https://shorensteincenter.org/digital-deceit-ii-policy-agenda-fight-disinformation-internet (accessed 8 Jun. 2019).
[68] Dommett, K. and Power, S., (2019), 'The Political Economy of Facebook Advertising: Election Spending, Regulation and Targeting Online', *The Political Quarterly*, 90(2): p. 260, doi: 10.1111/1467-923X.12687 (accessed 11 Sept. 2019).
[69] Microtargeting is essentially a more refined form of digital targeting.
[70] Vote Leave and the Labour Party are both reported to have used targeted advertising, as is the Trump campaign in the US. See Bartlett, Smith and Acton (2018), *The Future of Political Campaigning*.
[71] Manthorpe, R. (2019), 'Boris Johnson team posts hundreds of Facebook ads to test campaign messages', Sky News, 26 July 2019, https://news.sky.com/story/boris-johnson-team-posts-hundreds-of-facebook-ads-to-test-campaign-messages-11770644 (accessed 26 Jul. 2019).
[72] John, B. and Donno, C. (2019), 'UK election: How political parties are targeting voters on Facebook, Google and Snapchat ads', *First Draft*, 14 November 2019, https://firstdraftnews.org/latest/uk-election-how-political-parties-are-targeting-voters-on-facebook-google-and-snapchat-ads (accessed 25 Nov. 2019).
[73] Guru-Murthy, K. (2019), 'Target voter: how the parties are targeting you online during the election', *Channel 4*, 5 November 2019, https://www.channel4.com/news/target-voter-how-the-parties-are-targeting-you-online-during-the-election (accessed 25 Nov. 2019).
[74] Guardian Readers (2019), 'Have you seen any UK political adverts without disclosures on social media', *Guardian*, 5 November 2019, https://www.theguardian.com/media/2019/nov/05/have-you-seen-any-uk-political-adverts-without-disclosures-on-social-media-general-election (accessed 25 Nov. 2019).

can be optimized are time, location and content, while developments in the field of affective computing suggest that systems will be able to adapt to the mood of users and improve in sophistication.[75]

It has been suggested that microtargeting 'undermines the public sphere by thwarting public deliberation, aggravating political polarization, and facilitating the spread of misinformation'.[76] As former MEP Julia Reda has commented, voters have the right to know what a political party stands for[77] instead of being micromanaged.

While some aspects of targeting raise concerns, some researchers doubt that targeting with political intentions can be effective. Researcher Reuben Binns, for example, believes that apart from the difficulty of designing the right intervention, ML is not sufficiently advanced to establish an individual's 'ground truth'. Binns has stated: 'The situation is changing so rapidly, for ML to be effective you need to go beyond just the one point in time in the data set you have collected. To make something as generalizable and as robust over time, it has to be eventually latching onto causal mechanisms in the mind or in the social world, and we don't really have many ways to do that with ML.'[78]

The effectiveness of behavioural targeting based on psychographic modelling remains contested,[79] but companies specializing in consumer and media insights are willing to invest in the field,[80] while the personalization services industry is booming.[81] Commercial ad targeting is becoming the norm in digital advertising, with companies targeting people with the propensity to convert or suggesting that costumers' cognitive biases can be used to their clients' advantage. In an indication that the trend is likely to extend to legacy media institutions, the US publishing conglomerate Condé Nast[82] also offers a commercial targeted advertising solution via its Spire[83] platform. Sky has also rolled out its AdSmart service across Europe, offering audience segmentation for its advertising clients; and the UK's Channel 4[84] and Virgin Media[85] have also joined the AdSmart platform.

While media companies are exploring the commercial opportunities of personalized advertising, targeted political advertising remains a highly contested 'red line'. Paul Nemitz, the former director

[75] The European Commission-appointed High-Level Expert Group on Artificial Intelligence argued that 'Individuals should not be subject to unjustified personal, physical or mental tracking or identification, profiling and nudging through AI powered methods of biometric recognition such as: emotional tracking, empathic media, DNA, iris and behavioural identification, affect recognition, voice and facial recognition and the recognition of micro-expressions.' Nevertheless, affective AI is a field that marketers and retail companies are investing in already. see Gillespie, E. (2019), 'Are you being scanned? How facial recognition technology follows you, even as you shop', *Guardian*, https://www.theguardian.com/technology/2019/feb/24/are-you-being-scanned-how-facial-recognition-technology-follows-you-even-as-you-shop (accessed 22 Jul. 2019).
[76] Gorton (2016), 'Manipulating Citizens: How Political Campaigns' Use of Behavioral Social Science Harms Democracy', p. 61.
[77] Reda, J. (2019), 'Panel: Data is (political) power!', CPDP 2019 Conference, Brussels, 30 January 2019, https://www.youtube.com/watch?v=TopOHzLx3CE (accessed 8 Jul. 2019).
[78] Interview with the author, Chatham House, 24 January 2019.
[79] Matz et al. have reported successfully influencing the consumption choices of a group of users, but the jury is still out in terms of whether politically motivated psychological persuasion can be achieved by means of targeting. See Matz, S. C., Kosinksi, M., Nave, G. and Stillwell, D. J. (2017), 'Psychological targeting as an effective approach to digital mass persuasion', *Proceedings of the National Academy of Sciences* 114(48): pp. 12714–19, doi: 10.1073/pnas.1710966114 (accessed 11 Sept. 2019).
[80] In 2015, for example, Nielsen acquired Boston-based neuroscience firm Innerscope Research (subsequently renaming it as Nielsen Consumer Neuroscience).
[81] SmarterHQ, AB Tasty and Optimizely are just few of the companies selling personalization solutions.
[82] The company's print and online magazine titles include *The New Yorker*, *Wired*, *Vogue*, *GQ* and *Pitchfork*.
[83] Willens, M. (2019), 'Condé Nast launches new ad program for performance marketers', Digiday, 23 May 2019, https://digiday.com/media/conde-nast-launches-new-ad-program-performance-marketers (accessed 25 Jun. 2019).
[84] Channel 4 press office (2019), 'Sky and Channel 4 broaden industry-leading partnership', *Channel 4*, 17 September 2019, https://www.channel4.com/press/news/sky-and-channel-4-broaden-industry-leading-partnership (accessed 17 Nov. 2019).
[85] Kobie, N. (2019), 'The creepy world of personalised ads is coming to your TV', *Wired UK*, 1 October 2019, https://www.wired.co.uk/article/personalisted-tv-adverts-sky-adsmart-channel-4 (accessed 25 Nov. 2019).

responsible for fundamental rights and EU citizenship at the European Commission's DG Justice, has stressed that access to data relating to individuals' political opinions should be barred:

> Under GDPR parties have no right to know the individual voting intentions and to profile people for this purpose. Let's remember European history and dictators – the first they wanted to know was an individual's political opinion.[86]

The broader regulatory landscape is still in flux. In the UK, for example, the Data Protection Act 2018 allows registered political parties to process what is classified as 'special category' personal data revealing political opinions under an extended notion of 'substantial public interest'.[87] Until recently the US was the only market where Google allowed targeting based on political affiliation,[88] but in November 2019 the firm announced it would start imposing restrictions on the attributes it allowed targeting to be based on, excluding political affiliation or public voting records. Age, gender and postcode-level location would still be allowed, with the new policy pledged to be implemented ahead of the December 2019 UK general election.[89] Google itself prohibits targeting on the basis of race, religion, ethnicity and sexual orientation, among others.[90] In October 2019, in the context of increasing scrutiny of platforms from the US Congress and the EU, Twitter announced it would ban political advertising.[91] Twitter will continue to enable targeting for advertisements deemed non-political, based on behaviour signals, device, followers, keywords, geolocation and more.[92] Facebook allows targeting based on location, demographics, interests, behaviour and connections, among other signals, while its ongoing stance towards political ad targeting remains to be seen. Momentum has been building, however, so there is the prospect that the company will feel compelled to impose stricter restrictions on political ads sooner rather than later.

It is also important to note that the wider advertising technology (adtech) sector and programmatic advertising's real-time bidding (RTB) technologies have also come under scrutiny, notably by the UK's data protection regulator, the Information Commissioner's Office (ICO).[93]

[86] Interview with the author, 30 January 2019, CPDP Conference, Brussels.
[87] UK Data Protection Act 1998, Schedule 1, Part 2, Paragraph 22, http://www.legislation.gov.uk/ukpga/2018/12/schedule/1.
[88] Interview with Jon Steinberg, Senior Public Policy Manager at Google, Chatham House, 7 February 2019.
[89] Spencer, S. (2019), 'An update on our political ads policy', *Google – The Keyword*, 20 November 2019, https://blog.google/technology/ads/update-our-political-ads-policy (accessed 25 Nov. 2019).
[90] Google Inc. (2019), 'Advertising Policies Help: Personalised Advertising', https://support.google.com/adspolicy/answer/143465#papolicy.
[91] For more on social media's approach to political ads, see Culliford, E. (2019), 'Factbox: How social media sites handle political ads', Reuters, 15 November 2019, https://www.reuters.com/article/us-usa-election-advertising-factbox/factbox-how-social-media-sites-handle-political-ads-idUSKBN1XP22G (accessed 17 Nov. 2019).
[92] Dorsey, J. (@jack) (2019), 'We've made the decision to stop all political advertising on Twitter globally. We believe political message reach should be earned, not bought. Why? A few reasons…', tweets, 30 Oct. 2019, https://twitter.com/jack/status/1189634360472829952 (accessed 26 Nov. 2019). For more information, see https://business.twitter.com/en/targeting/behavior.html.
[93] The ICO has launched an investigation into adtech, in the context of which it hopes to cooperate with other data protection authorities. See ICO (2019), *Update report into adtech and real time bidding*, 20 June 2019, https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf (accessed 3 Jul. 2019).

4. Legacy Media and Personalization

> I'm deeply concerned that we don't have the foggiest clue how to approach the media landscape today. – danah boyd, SXSW EDU 2018.[94]

In terms of personalization, the majority of legacy media are currently testing or launching recommender systems, with larger conglomerates also incorporating targeted advertising. Multinational media networks such as Viacom (the parent company of Comedy Central, Paramount Network and Nickelodeon, among others), are harnessing terabytes of log files, first- and third-party data on a monthly basis, to optimize their content and conduct cross-platform personalization.[95] There is much scope for employing its digital recommender systems, as even though Viacom's current digital household reach is still a small portion of its linear audience (i.e. that views content at the time of broadcast, such as on 'traditional' TV), the average time digital-only households spend using Viacom's services tends to be markedly higher than linear households.[96] NBC's digital division is also employing recommendation engines using collaborative filtering,[97] as is Vox Media.[98]

In December 2018 Meredith Kopit Levien, Chief Operating Officer at the *New York Times,*[99] announced that the company was to invest heavily in AI and ML to support personalization, hoping it could use 'an enormous amount of active and passive signals from people', to eventually show them what is 'more interesting to them'.[100] The *New York Times* uses 'multi-armed bandit' models[101] to address the problem of 'exploration versus exploitation'[102] and, in contrast to other companies, it is not using third-party data to train its recommendation systems. More specifically, its algorithms take a self-adjusting 'contextual-bandit' approach,[103] that operates according to a time-decay principle (weighting recent data more highly than older signals), and retrains every 15 minutes. This approach allows for more serendipity into the system.[104]

Recommendations are just one aspect of the content personalization employed by legacy media. Tailoring can be applied to each of the following: the timing of content delivery; headlines; newsletters; section visibility and interface layout; social media posts; geolocated customization;

[94] boyd, d. (2018), 'What Hath We Wrought?', SXSW EDU 2018 keynote speech, https://www.sxswedu.com/news/2018/watch-danah-boyd-keynote-what-hath-we-wrought-video (accessed 25 Jun. 2019).
[95] Fogelson, S. and Lipson, S. (2019), 'Viacom's nique Audience Platform: Unifying our Audiences to Deliver an Even Better Viacom Experience', Presentation, Data Science Salon New York, 13 June 2019.
[96] Ibid.
[97] Edwards, S. (2018), 'The Beauty and Elegance of Collaborative Filtering and its Surprising Utility in Producing Useful Streaming Video Recommendations', Presentation, Data Science Salon Los Angeles, 13 September 2018.
[98] Vox Media is the parent company of digital natives Vox, The Verge, and Recode, among others.
[99] Open, the company's tech blog, is useful for regular updates: https://open.nytimes.com.
[100] Slefo, G. P. (2018), 'New York Times plans to invest heavily in AI to improve personalization', Ad Age, 3 December 2018, adage.com/article/digital/york-times-poised-copy-facebook/315831 (accessed 22 May 2019).
[101] Coehen, A. (2019), 'Algorithmic Recommendations at the New York Times', Presentation, Data Science Salon, New York, 13 June 2019. For a thorough treatment of multi-armed bandits, see Slivkins, A. (2019), 'Introduction to Multi-Armed Bandits', Microsoft Research NYC, April 2019, https://arxiv.org/abs/1904.07272 (accessed 25 Jun. 2019).
[102] The tension between exploitation and exploration relates to the fact that by selecting a specific model predicted to lead to high CTR for example, comes at the cost of exploring different models with uncertain but also potentially better results.
[103] The *New York Times*'s models are adapted from the 2010 paper of Li, L., Chu, W., Langford, J. and Schapire, R. E., entitled 'A Contextual-Bandit Approach to Personalized News Article Recommendation' and presented at the 19th International Conference on World Wide Web (WWW 2010), Raleigh, NC, https://arxiv.org/abs/1003.0146.
[104] For more information on how the *New York Times* employs personalization see https://help.nytimes.com/hc/en-us/articles/360003965994-Personalization (accessed 18 Nov. 2019)

and push notifications, as well as to the content itself.[105] For example, John Keefe, currently Investigations Editor at digital-native publisher Quartz, has been looking into conversational interfaces and personalization in push notifications.[106] The *New York Times* has also launched a flash briefing, available through Alexa-enabled devices, that has the potential to become more personalized. The *Washington Post* segments audiences based on age (so-called millennials, for example), influencer status, politics (left-leaning, etc.),[107] and has built a series of digital products – such as Clavis or Bandito – to optimize its services, some of which it licenses to other publishers and broadcasters through its Arc Publishing[108] platform. The *Washington Post* recorded 86.6 million unique visitors in March 2019, marking a 5.5 per cent month-on-month increase.[109]

The UK's public broadcaster, the BBC, offers personalization and recommendations for a variety of its services,[110] but much of its activity in the field of personalization remains a work in progress, testing collaborative and semantic filtering and content personalization. Signed-in BBC account holders can set their location to receive local news and weather reports. The *Guardian* offers the opportunity for users to adjust their homepage depending on their location (options include the US, Australia and the UK), but its strong subscriber base can provide it with data for efficient personalization. At *The Times* and the *Sunday Times*, owned by News UK, the trialling in 2018 over a nine-month period of JAMES, a new software using different ML algorithms to produce personalized newsletters, led to a 49 per cent drop in the number of digital subscription cancellations.[111] Dan Gilbert, director of data technology at News UK, stated in an interview with the author that the company is still grappling with the balance between algorithmic modelling and human control.[112]

Switzerland's biggest private media group, Tamedia, also uses personalization (notably, by means of its textbot, known as 'Tobi'),[113] while Norway's Schibsted media firm claims to see it as a way of closing of the gap between 'what people know and what they should know'.[114] Schibsted reaches around 80 per cent of audiences in both Norway and Sweden.[115] Finnish public broadcaster Yle is also using personalization extensively, and employs collaborative filtering for streamed videos and article recommendations.[116]

Since 2017 the European Commission has invested in personalization by providing almost €4 million in funding, through its Horizon 2020 research and innovation programme, to Content

[105] Plattner, T. (2018), 'Ten effective ways to personalize news platforms', Medium, 14 April 2018, https://medium.com/jsk-class-of-2018/ten-effective-ways-to-personalize-news-platform-c0e39890170e (accessed on 8 Jun. 2019).
[106] Interview with the author, 28 February 2019.
[107] Prakash, S. (2017), 'Journalism and technology: Big data, personalization and automation', Keynote address, Computation + Journalism Symposium, October 2017, Northwestern University, Evanston, IL, https://www.youtube.com/watch?v=PqMvx089AQ4 (accessed 6. Jul. 2019).
[108] Arc Publishing, https://www.arcpublishing.com (25 Nov. 2019).
[109] Washington Post PR (2019), 'The Washington Post records 86.6 million unique visitors in March 2019', *The Washington Post*, 17 April 2019, https://www.washingtonpost.com/pr/2019/04/17/washington-post-records-million-unique-visitors-march (accessed on 17 Nov. 2019).
[110] Signed-in users can personalize the BBC Homepage, and access recommendations for the BBC iPlayer or the BBC+ app, for example.
[111] Tobitt, C. (2019), 'Times titles halve digital subscriber churn with tailored emails from AI named "James", Press Gazette, 27 May 2019, https://www.pressgazette.co.uk/times-titles-halve-digital-subscriber-churn-with-tailored-emails-from-ai-named-james (accessed 4 Jun. 2019).
[112] Interview with the author at News UK, 5 April 2019.
[113] See Plattner, T. and Orel, D. (2019), 'Addressing Micro-Audiences at Scale', Computation + Journalism Symposium, February 2019, Miami.
[114] Diakopoulos, N. (2019), *Automating the News: How the Algorithms are Rewriting the Media*, Cambridge, MA: Harvard University Press, p. 195.
[115] Schibsted Media Group (2018), 'Schibsted will be divided into two companies', 18 September 2018, https://schibsted.com/news/schibsted-will-be-divided-into-two-companies (accessed 30 Jun. 2019).
[116] Interview with Jaakko Lempinen, head of customer experience, data and AI at Yle, 5 June 2019.

Personalization Network (CPN), a consortium of media,[117] university and technical partners that aim to create a recommendation tool. In an effort to avoid so-called 'filter bubbles' – a term used to describe a reduction in content diversity or in contrasting views – CPN injects serendipity into its models. Various aspects of CPN's work are public and the project invites audience feedback.[118]

Considerations for legacy media and their mission

Apart from the tension between scale and quality, personalized news recommendation systems have deep political effects, and 'the decision criteria embodied in their systems have profound political consequences' for democracy.[119] Algorithmic systems' core function is to predict, but 'the change in the standard for news selection from judgment to prediction [of what readers want] inevitably changes the character of the news [...]'.[120] Telling audiences what they want – or expect – to hear is markedly different from telling them what they need to hear. Positioning audiences' preferences at the centre of the journalistic endeavor, and thereby jeopardizing the autonomy of journalists, could diminish the latter's ability to serve the public interest.[121] The balance between serving such a broad client base as 'the public' while delivering individualized content is extremely fragile.

The digital marketing industry has also started pushing for personalization, with some companies producing reports calling on journalism to adopt the technology in order to save itself.[122] All the while, behaviour-driven adtech targeting can render legacy media titles less pertinent, by putting the focus on 'the person rather than the publication'.[123] The reframing of news as a commodity to be consumed by way of quantifiable clicks is in tension with the public-interest character of the Fourth Estate. More broadly, the idea of the internet as part of a public sphere does not translate easily to the reductive notion of the internet as a marketplace.[124]

Algorithmic optimization was created for e-commerce applications, and the question of whether its embedded goals serve the values of public-interest journalism remains unanswered. Nick Diakopoulos, an academic specializing in the study of algorithmic journalism, has argued that 'because of its affordances for scale and speed, automation creates a "more, more, more" mentality', but has also insisted that the ethical deployment of AI and ML in news production calls for measured consideration of whether more should mean 'more quality rather than more output'.[125]

[117] Three out of the nine participants in CPN are media partners – Deutsche Welle of Germany, VRT of Belgium and Dias Publishing of Cyprus – but CPN are searching for outside media partners to test their personalization tool, which is currently in a testing phase. See https://www.projectcpn.eu.
[118] A report on the platform's architecture and the technological infrastructure can be accessed here: https://static1.squarespace.com/static/595cd20e1b10e30e621770e9/t/5cd558a824a69427a294007b/1557485754120/D3.3+Technology+Bricks+V2.pdf.
[119] MacCarthy, M. (2019), 'The Ethical Character of Algorithms – and What It Means for Fairness, the Character of Decision-Making, and the Future of News', Shorenstein Center, 15 March 2019, https://ai.shorensteincenter.org/ideas/2019/1/14/the-ethical-character-of-algorithmsand-what-it-means-for-fairness-the-character-of-decision-making-and-the-future-of-news-yak6m (accessed 28 Jun. 2019).
[120] Ibid.
[121] Tandoc, Jr, E. C. and Thomas, R. J. (2015), 'The Ethics of Web Analytics: Implications of using audience metrics in news construction', *Digital Journalism*, 3(2): p. 244, doi: 10.1080/21670811.2014.909122 (accessed 11 Sept. 2019).
[122] Cxense (2019), *Don't stop the press! What does the future hold for journalism?*, https://cdn2.hubspot.net/hubfs/1945032/Premium-content/Cxense_Future_of_Journalism_2019.pdf (accessed 3 Jul. 2019).
[123] Bakir, V. and McStay, A. (2018), 'Fake News and the Economy of Emotions', *Digital Journalism*, 6(2): p. 166, doi: 10.1080/21670811.2017.1345645 (accessed 11 Sept. 2019).
[124] Peacock, S. E. (2019), 'How web tracking changes user agency in the age of Big Data: The used user', *Big Data & Society*, 1(2): p. 2, doi: 10.1177/2053951714564228 (accessed 11 Sept. 2019).
[125] Diakopoulos (2019), Automating the News: How the Algorithms are Rewriting the Media, p. 7.

One of the findings of the 2019 Reuters Institute Digital News Report, that audiences thought news media were doing 'a better job at breaking news than explaining it',[126] may be indicative of a pushback against prioritizing output over quality.

The issue of whether or not an algorithmically-driven model may be more efficient than an editorially-driven one can only be answered if targets are clearly defined, metrics appropriately selected and effects closely monitored. Targets dictate how the models are built. The semantics of data built into algorithmic models affect what surfaces – for example, long term trends (inequality) or peaks (extreme events) – thus, different journalistic approaches will need different models. Optimizing for wide-range analysis that reflects the complexities of an event is markedly different from optimizing with the purpose of keeping users informed about breaking news. The precise design not only ultimately determines content diversity, it can increase it.[127] The parameters used for optimization and the weightings of different features in the models are important in that regard. In any case, legacy media should be transparent about the data they collect, the way they process it and how they store it – research has indicated certain media outlets have not fully internalized the concept of proper data stewardship, and have been disclosing information to third parties without being transparent to users.[128]

AI and personalization, like other technologies that preceded them, are bound to have feedback effects on the journalistic profession, changing power dynamics between journalists and the commercial imperatives of media organizations, especially in markets with increasing consolidation. The domestic media environment is important, as the structure and business orientation of media organizations will dictate the metrics for which they will choose to optimize their systems.

A further side effect of the personalization drive that should not be overlooked is the fact that the surveillance apparatus which it is gradually putting in place will render anonymity of protected sources and whistle-blowers harder to defend.

[126] Newman, Fletcher, Kalogeropoulos and Nielsen (2019), *Reuters Institute Digital News Report 2019*, p. 10.
[127] Möller, Trilling, Helberger and van Es (2018), 'Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity'.
[128] Binns, R. and Libert, T. (2019), 'Good news for people who love bad news: centralization, privacy, and transparency on US news sites', WebSci '19, June 30–July 3, 2019, Boston, MA; and Harlow, M. and Murgia, M. (2019), 'How top health websites are sharing sensitive data with advertisers', *Financial Times*, 13 November 2019, https://www.ft.com/content/0fbf4d8e-022b-11ea-be59-e49b2a136b8d (accessed 19 Nov. 2019).

5. Broader Implications of Personalization

Implications for individuals

Personalization's implications for individual citizens can be manifold. Discrimination could become more difficult to detect, and privacy more easily compromised, via increasingly frictionless[129] methods. Freedom of opinion and expression, autonomy and agency may be thwarted by the development of filter bubbles and self-reinforcing polarized clusters, termed 'echo chambers'.

While the filter bubbles argument is not supported by a concisely robust body of evidence, it is important to be aware that this is an under-researched field, mainly because of highly controlled access to the data produced as a result of the economies of scale practised by 'big tech'. According to research from DeepMind, recommender systems can lead to filter bubbles or echo chambers, with the injection of random exploration and increases in the pool of candidate content serving as potential countermeasures.[130] As we have seen, various news recommender teams are adopting this approach.

Research challenging the existence of filter bubbles[131] has its own constraints, such as limited samples,[132] application-specific surveys that may provide an incremental view of the system, or the inherent complexity of accurately measuring exposure in the digital environment. New issues, such as the promotion of a limited set of actors that are already market-dominant in the information space, have also been raised.[133] Other research has cast doubt on the existence of filter bubbles, based on a lack of audience fragmentation at the market level of online and offline media consumption,[134] but personalization opens the door to a more granular fragmentation at the level of each individual news source.

One of the great challenges of research into personalization in the media is the fact that communication effects accumulate over time. Digital interactions that may seem trivial in platform- or time frame-specific analysis may be impactful in aggregate, especially in high-dimensional ecosystems[135] and through the deployment of ML models that are susceptible to hidden feedback loops.

[129] Facebook's announced Threads app is an example. See Hern, A. (2019), 'Facebook preparing new app to maintain pressure on Snapchat', *Guardian*, 27 August 2019, https://www.theguardian.com/technology/2019/aug/27/facebook-new-app-threads-instagram-pressure-on-snapchat (accessed 11 Sept. 2019).
[130] Jiang, R., Chiappa, S., Lattimore, T., György, A. and Kohli, P. (2019), 'Degenerate Feedback Loops in Recommender Systems', Proceedings of AAAI/ACM Conference on AI, Ethics, and Society, Honolulu, January 27–28, 2019 (AIES '19), https://arxiv.org/abs/1902.10730 (accessed 11 Sept. 2019).
[131] Axel Bruns, for example, rejects the concept in his book *Are Filter Bubbles Real?* (2019, Cambridge and Medford, MA: Polity Press), calling the surrounding debate a 'moral panic', but also makes a series of assumptions about users' unfettered agency, their use of a variety of platforms, and more.
[132] For example, Haim, M., Graefe, A. and Brosius, H.-B. and in their article 'Burst of the Filter Bubble?: Effects of personalization on the diversity of Google News' (2018, *Digital Journalism*, 6(3): pp. 330–43) monitored the explicit and implicit personalization results of the Google News aggregator for a couple of days; Möller, Trilling, Helberger and van Es in 'Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity' (2018), researched the recommendation system of just one Dutch newspaper.
[133] Nechushtai and Lewis (2019), 'What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations'.
[134] Fletcher, R. and Nielsen, R. K. (2017), 'Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication', *Journal of Communication*, 67(4): pp. 476–98.
[135] High-dimensional data relate to datasets whose inputs can have a high number of dimensions or attributes.

It is important not to lose sight of what Selbst et al. have termed the 'ripple effect trap': 'the failure to understand how the insertion of technology into an existing social system changes the behaviors and embedded values of the pre-existing system'.[136] Some have suggested examining these phenomena while being attentive to the fact that individuals are often positioned at the intersection of 'technology, culture and class', [137] while others have outlined how radicalization can take place in the fringes of online discourse.[138]

Freedom of opinion and expression

A report by David Kaye, the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, warned that personalization that minimizes the diversity of news citizens access can 'reinforce biases', while optimizing for engagement may undermine users' ability to seek and find information.[139]

Autonomy and agency

UN Special Rapporteur Kaye also called for 'rights-oriented research into the social, economic and political effects of AI-assisted curation', warning AI's opacity risks interfering with 'individual autonomy and agency'.[140] Moreover, he called for human rights audits with respect to AI's automation function, the data analysis that feeds into the models and its adaptability. The Council of Europe has also stated that 'fine grained, sub-conscious and personalized levels of algorithmic persuasion may have significant effects on the cognitive autonomy of individuals and their right to form opinions and take independent decisions'.[141] Autonomy should not be narrowly construed merely as the individuals' ability to make decisions, but as the freedom to 'deliberate and act as partially dependent on the myriad material and social relationships in which they are situated'.[142]

Algorithmic systems could make it difficult to distinguish between offering, persuasion and manipulation.[143] As Sunstein has stated: 'Choice-making is a muscle',[144] and consistently personalized offerings may result in atrophy. AI-driven personalization could threaten citizens' autonomy – their ability to govern themselves based on their own desires, characteristics and circumstances. Each of us is vulnerable in a certain way – now our vulnerabilities are easier to detect, through our active consumer inputs or through algorithmic inferences made about us. AI-derived inferences, often used to close gaps in datasets, can compromise the accuracy of the deployed models.

[136] Selbst, A. D., boyd, d., Friedler, S. A., Venkatasubramanian, S. and Vertesi, J. (2018), 'Fairness and Abstraction in Sociotechnical Systems', 23 August 2018, ACM Conference on Fairness, Accountability, and Transparency (FAT*), 1(1): p. 6, https://ssrn.com/abstract=3265913 (accessed 11 Sept. 2019).
[137] Davies, H. C. (2018), 'Redefining Filter Bubbles as (Escapable) Socio-Technical Recursion', *Sociological Research Online*, 23(2): p. 3.
[138] Kaiser, J. and Rauchfleisch, A. (2019), 'Integrating Concepts of Counterpublics into Generalised Public Sphere Frameworks: Contemporary Transformations in Radical Forms', *Javnost – The Public*, 26(3), doi: 10.1080/13183222.2018.1558676 (accessed 11 Sept. 2019).
[139] United Nations (2018), Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 29 August 2018, A/73/348.
[140] Ibid.
[141] Council of Europe – Committee of Ministers (2019), *Declaration by the Committee of Ministers on the Manipulative Capabilities of Algorithmic Processes*, 13 February 2019, https://search.coe.int/cm/pages/result_details.aspx?ObjectId=090000168092dd4b.
[142] Owens, J., Mladenov, T. and Cribb, A. (2017), 'What Justice, What Autonomy? The Ethical Constraints upon Personalisation', *Ethics and Social Welfare*, 11(1): p. 13, doi: 10.1080/17496535.2016.1234631 (accessed 11 Sept. 2019).
[143] Cave, S. (2017), *Written Evidence For the House of Lords Select Committee on Artificial Intelligence*, http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69702.html (accessed 8 Jul. 2019).
[144] Sunstein, C. (2015), 'The Ethics of Nudging', Yale Journal on Regulation, 32(2): p. 436.

Data protection

The Charter of Fundamental Rights of the EU enshrines data protection as a fundamental right for everyone to enjoy control over access to their data and its rectification. The consent of social and legacy media audiences for their data to be used in algorithmic models may not be transparent, while researchers have also highlighted the fact that algorithmic models may be interrogated by model inversion attacks to glean training data, leading some to argue that some ML models can be classified as 'personal data in their own right'.[145]

Dignity and isolation

The issue of dignity – the notion of being respected in the context of interpersonal connections – may not be discussed very often, but its loss, coupled with social isolation, can be detrimental to the individual, as it is embedded in his/her sense of self. Deploying personalization in healthcare contexts – and specifically in the field of mental health[146] – poses its own risks in terms of preserving patients' dignity. Additionally, European conceptions of privacy are grounded on the concept of dignity, so the current challenges to the former have resonances for the latter. The concept of dignity is relational, so personalization's fragmentation of multinodal channels of communication is bound to undermine it by enhancing isolationist tendencies.

Discrimination

Discrimination is inbuilt within every algorithmic system. As the New York-based AI Now Institute has emphasized, 'AI systems function as systems of discrimination: they are classification technologies that differentiate, rank, and categorize.'[147] Personalization can undermine people's ability to monitor wrongful discrimination, and when it takes the form of exclusion by filtering out and silencing minority voices, democracy-preserving social change may be obstructed, as these groups are often drivers of change. Discrimination can take place not just on the basis of personal data or legally protected characteristics, but on assumed, inferred interests, this being what Wachter has called 'discrimination by association'[148] with a group. Near-perfect price discrimination, with companies charging the highest fee that a customer would be willing to pay, will be possible too[149] if consumer protection laws are not updated in time. New patents, such as Facebook's purported facial recognition system to match customers in bricks-and-mortar stores with user profiles,[150] or scan photos for products,[151] can exacerbate this risk.

[145] Veale, M. (2019), Doctoral Thesis: 'Governing Machine Learning that Matters', University College London – Department of Science, Technology, Engineering and Public Policy, p. 138.
[146] For more on AI deployment in mental healthcare, see Burr, C., Morley, J., Taddeo, M., and Floridi, L. (2019), 'Digital psychiatry: ethical risks and opportunities for public health and well-being', October 30, 2019. Available at SSRN: https://ssrn.com/abstract=3477978 or http://dx.doi.org/10.2139/ssrn.3477978
[147] West, S. M., Whittaker, M. and Crawford, K. (2019), *Discriminating Systems: Gender, Race and Power in AI*, New York: AI Now Institute, https://ainowinstitute.org/discriminatingsystems.pdf (accessed 30 Jun. 2019).
[148] Wachter, S. (2019), 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising', https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3388639, (accessed 11 Sept. 2019).
[149] Ezrachi, A. and Stucke, M. E. (2016), 'The rise of behavioural discrimination', *European Competition Law Review*, 37(12): pp. 485–92.
[150] Singer, N. (2018), 'Facebook's push for facial recognition prompts privacy alarms', *The New York Times*, 9 July 2018, https://www.nytimes.com/2018/07/09/technology/facebook-facial-recognition-privacy.html (accessed 24 Jul. 2019).
[151] Locker, M. (2019), 'Creepy Facebook patent uses image recognition to scan your personal photos for brands', *Fast Company*, 4 October 2019. https://www.fastcompany.com/90333067/creepy-facebook-patent-uses-image-recognition-to-scan-your-personal-photos-for-brands (accessed 17 Nov. 2019).

Identity

ML algorithms can be characterized by the so-called *stationary assumption*, as they tend to model behaviours assumed to be consistent through time. Hence ML-driven personalization can also have implications for identity, as it tends to promote a static conception of personhood[152] despite the fact that individuals adopt different identities to navigate and assist their social interactions and their personal development. Individuals' goals and identities are in flux, especially in this interconnected world. Pariser has also argued that filter bubbles can disconnect us from our ideal selves – that which we aspire to be – and personalization 'overfitting'[153] can misdiagnose our behaviours. Algorithmic systems built on old training data while serving personalities that are meant to be evolving can lead to *concept drift*[154] and so become not fit for purpose. In response to this issue, 'drift-aware' adaptive algorithms have started to be developed.[155]

Privacy

AI's ability to infer intimate data from the available datasets poses serious privacy concerns. When the models are relying on and interpreting unconscious processes, not just the origins but the very existence of manipulation might go undetected by the targeted users. 'Trap design' tends to blur the boundaries between persuasive and coercive strategies, while the conflation of user retention (measured by what Seaver has called captivation metrics) and satisfaction should not be taken at face value.[156]

Profiling[157]

According to Article 22 of the EU's General Data Protection Regulation (GDPR),[158] individuals have the right not to be subject to profiling. However, a lack of transparency prevails as to *how* citizens are being profiled and targeted. The implications of cross-device and cross-platform tracking, segmenting and profiling are manifold. Fairness, accountability, bias, discrimination or exploitation of vulnerable groups (including children) are key issues that need to be addressed in personalized communication, especially in relation to personalized persuasive technologies.[159] Specifically with regard to children, the European Commission-appointed High-Level Expert Group (HLEG) on Artificial Intelligence has called for close monitoring of personalized systems built on children's profiles, and has even suggested a consideration of whether children should receive a 'clean data

[152] Pariser, E. (2011), The Filter Bubble: How The New Personalized Web Is Changing What We Read and How We Think, Penguin Books, p. 218.
[153] Ibid., p. 130. Overfitting in ML is the mistake of identifying patterns that are not actually there, by memorizing training data rather than detecting generalizable patterns.
[154] Concept drift relates to the situation where ML systems using static models based on historical data become prone to errors, as the real-life context in which they operate changes along with the relevant data.
[155] Gaba, J. Žliobaitė, I., Bifet, A., Pechenizkiy, M. (2019), 'A survey on concept drift adaptation', *ACM Computing Surveys (CSUR)*, 4(4), April 2014
[156] Seaver, N. (2019), 'Captivating algorithms: Recommender systems as traps', *Journal of Material Culture*, 24(4): pp. 9–11, https://doi.org/10.1177/1359183518820366 (accessed 11 Sept. 2019).
[157] Data brokers' core business model, typified by those of Acxiom and Oracle, is based on profiling. In November 2018 Privacy International filed complaints against the two data brokers, but also against the adtech companies Criteo, Quantcast and Tapad, and the credit referencing agencies Equifax and Experian.
[158] See GDPR, Article 22: 'Automated individual decision-making, including profiling', https://gdpr-info.eu/art-22-gdpr (accessed 6 Jul. 2019).
[159] Kaptein, M., Markopoulos, P., de Ruyter, B. and Aarts, E. (2015), 'Personalizing persuasive technologies: explicit and implicit personalization using persuasion profiles', *International Journal of Human-Computer Studies*, 77: pp. 38–51, doi: 10.1016/j.ijhcs.2015.01.004 (accessed 22 May 2019).

slate' of the data related to their childhood.[160] Both the UK and the US have taken steps towards protecting children from profiling: in the case of the UK, through the ICO's draft code of practice on age-appropriate design; and in the US, through the Federal Trade Commission's Children's Online Privacy Protection Act (COPPA). It is estimated that one in three current internet users are children.[161]

Societal and political implications

The network effects of digital intermediaries, and the technologies to which they have given rise, mean that any externalities will also scale, with ripple effects on wider society and states. Effects can cascade through layers, from the individual to group, society and national level. Notwithstanding the crucial importance of individual rights, collective rights must be protected too.

Political security

Political security (which in this paper, as noted earlier, concerns the organizational stability of states, and the ideologies that provide them with legitimacy) relates to social cohesion and/or polarization, but at the same time is broader than that. The political stability and legitimacy of individual states may be undermined by uninhibited and unchecked personalization. Current personalization efforts remain predominantly a trial-and-error process, and empirical evidence of filter bubbles is lacking, but if personalization improves and its reach scales up, issues could arise for democracy and its capacity to be deliberative and reflective through an informed citizenry.[162] The fear of filter bubbles is not going to subside in the foreseeable future, but in terms of mitigating steps there is no one-size-fits-all solution. Different conceptions of democracy and social contract might be affected in different ways, with some prioritizing public deliberation while others give priority to autonomy, plurality or freedom of choice.[163] In any scenario, the right to receive information is fundamental for political participation. Algorithmically-driven personalization systems can undermine 'the fairness and the quality of political discourse',[164] by complicating the free exchange of ideas.

Social cohesion

Eskens et al. have argued that states have positive obligations in relation to personalized news, in the light of Article 10 of the European Convention on Human Rights (ECHR) on the right to freedom of expression and information, and have pointed to 'exposure diversity' – people being exposed to truly diverse information – as an enhancer of social cohesion.[165] Personalized

[160] European Commission: Independent High-Level Expert Group on Artificial Intelligence (2019), *Policy and investment recommendations for trustworthy Artificial Intelligence*, European Commission, DG Connect, 26 June 2019, p. 40, https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence (accessed 11 Sept. 2019).
[161] Livingstone, S., Carr, J. and Byrne, J. (2016), *One in Three: Internet Governance and Children's Rights*, Innocenti Discussion Paper No. 2016-01, Florence: UNICEF Office for Research, https://www.unicef-irc.org/publications/pdf/idp_2016_01.pdf (accessed 11 Sept. 2019).
[162] Zuiderveen Borgesius et al. (2016), 'Should we worry about filter bubbles?', p. 10.
[163] Bozdag, E. and van den Hoven, J. (2015), 'Breaking the filter bubble: democracy and design', *Ethics and Information Technology*, 17(4): pp.249–65.
[164] Mittelstadt, B. (2016), 'Auditing for Transparency in Content Personalization Systems', *International Journal of Communication*, 10, https://ijoc.org/index.php/ijoc/article/view/6267/1808, p. 4992.
[165] Eskens, S. J., Helberger, N. and Möller, J. E. (2017), 'Challenged by news personalisation: five perspectives on the right to receive information', *Journal of Media Law*, 9(2): pp. 259–84, doi: 10.1080/17577632.2017.1387353 (accessed 11 Sept. 2019).

communication may reshape political campaigning into addressing citizens as single-issue voters, targeted only with the policy issue determined to be more affective on them. Clustering voters into a single issue facilitates their easier control and engagement – the division of the UK electorate into pro- and anti-Brexit constituents certainly enables certain actors, while disempowering others. As Richard Semiatin remarked: 'increasingly, campaigns will become about *you*, the voter, or [...] *you*, the consumer'.[166] This can have implications for social cohesion more broadly, by undermining community solidarity.[167] Citizens' capacity to jointly debate, develop and draw on their collective intelligence in order to meaningfully address issues might be diminished in a fragmented public sphere no longer able to recalibrate individual incentives to reach consensus.

Disinformation

The use of personalized messages by Russia's Internet Research Agency to target segments of the US population with customized disinformation via Facebook and Instagram has been widely reported,[168] and is symptomatic of the increasing individualization of citizenry to achieve campaign goals. By addressing the individual, the mass disinformation campaigns managed to evade public oversight. Social media companies have since launched ad libraries indexing political advertisements, the efficiency of which has been criticized severely,[169] while other policy changes may counteract any gains made by these transparency reports.[170] In any case, it is highly unlikely that social media measures without broader electoral and media reform will contain the issue.

Polarization

AI-driven targeting and personalization have also been blamed for the public sphere's increasing polarization by Ben Scott, senior adviser at the New America think-tank and policy director at Luminate, during his testimony to the International Grand Committee on Big Data, Privacy and Democracy meeting in Ottawa in May 2019.[171] In June, Tristan Harris, co-founder and executive director at the non-profit Center for Humane Technology, told the US Senate's Subcommittee on Communication, Technology, Innovation, and the Internet that 'the polarization of our society is [...] part of the business model' of internet platforms.[172] It is well established by now that emotional content drives engagement, and when the latter is the basis of digital platforms' incentive structures

[166] Semiatin, R. J. (2012), 'Introduction – Campaigns on the Cutting Edge', in Semiatin, R. J. (ed.), *Campaigns on the Cutting Edge* (2nd edn), Thousand Oaks, CA: CQ Press, p. 4.
[167] Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., and Cave, S. (2019), *Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research*. London: Nuffield Foundation, p. 20.
[168] Howard, P. N., Ganesh, B., Liotsou, D., Kelly, J. and François, C. (2018), *The IRA, Social Media and Political Polarization in the United States, 2012–2018*, Computational Propaganda Research Project, University of Oxford, https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/IRA-Report.pdf (accessed 8 Jul. 2019).
[169] Rosenberg, M. (2019), 'Ad tool Facebook built to fight disinformation doesn't work as advertised', *New York Times*, 25 July 2019, https://www.nytimes.com/2019/07/25/technology/facebook-ad-library.html (accessed 26 Jul. 2019).
[170] Sunstein, C. R. (2019), 'Facebook Can Fight Lies in Political Ads', Bloomberg, 9 October 2019, https://www.bloomberg.com/opinion/articles/2019-10-09/facebook-can-fight-lies-in-political-ads (accessed 14 Oct. 2019).
[171] International Grand Committee Hearing on Big Data, Privacy, and Democracy: Ottawa, Canada. YouTube, 28 May 2019, https://www.youtube.com/watch?time_continue=89&v=5fv105SGnd4&feature=emb_logo (accessed 17 Nov. 2019).
[172] US Senate Committee on Commerce, Science and Transportation: Subcommittee on Communication, Technology, Innovation and the Internet (2019), *Hearing: Optimizing for Engagement: Understanding the Use of Persuasive Technology on Internet Platforms*, 25 June 2019, https://www.commerce.senate.gov/public/index.cfm/2019/6/optimizing-for-engagement-understanding-the-use-of-persuasive-technology-on-internet-platforms (accessed 30 Jun. 2019).

it is difficult not to take seriously Harris's comments. In the long term, the social sorting[173] of populations by algorithmic classification and management can exacerbate polarizing tendencies.

The opportunities of personalization

Beyond the profits of actors occupying competitive positions in the digital marketplace, such as social media, digital retailers, adtech firms and data brokers, ML-driven personalization in the information space has not produced enough benefits to counteract its risks. The balance between an attention economy that profits digital oligopolies and one that has equally visible benefits for the individual and society is skewed to a concerning degree.

But AI deployment forces a day of reckoning when, across all sectors, vulnerabilities, inequalities and inefficiencies that have been allowed to perpetuate through political short-termism or difficulties in galvanizing collective action will need to be addressed decisively – that is, if the global crises that can be envisaged as systems entrench, scale and automate existing problems, are to be averted. It is crucial, now more than ever, to debate and commit to democracy's overarching goals and pursue targeted personalization policies that can achieve them. ML-personalized systems of communication can increase meaningful civic engagement[174] and assist consensus-building capacities by efficiently coordinating dialogue between informed citizens. Personalized communication can be employed to alert citizens to events and policies they deem important, or to fight the dissemination of disinformation by targeting debunks.

AI may be used to glean the consensus on certain issues,[175] but that does not mean it can or should substitute what Fourth Estate journalism offers. Context-aware recommender systems (CARS) can provide contextual conflict reporting that puts long-term confrontations into perspective.[176] Personalized recommendations also provide legacy media with the ability to increase the lifespan of 'evergreen' articles,[177] and to serve their democracy-preserving role by better reflecting the diversity of their audience and thereby supporting inclusivity. Marginalized and minority voices that felt they were formally excluded from 'traditional' media coverage, not least because of limited space, could now be represented more prominently in the online communities to which they relate or the broader audiences they need to reach. The exposure afforded to marginalized groups should be commensurate to their support for democracy, as extremist and dangerous elements should not be amplified just by virtue of occupying the periphery of public discourse. For the trade-offs between the risks and opportunities of personalization to be properly assessed, it will be critical to instil more clarity in terms of its scope and limitations.

[173] Lyon, D. (ed.) (2003), Surveillance as Social Sorting: Privacy, risk, and digital discrimination, New York: Routledge.
[174] Citizen participation platform Consul, for example, is working to employ ML to assist citizens' collaboration on putting forward mostly local-level proposals. See http://consulproject.org/en.
[175] Google News founder Krishna Bharat, speaking at the Global Editors Network (GEN) Summit 2019, argued that AI can also totally transform journalism by providing real-time fact-checking to interviewers, going as far as to suggest that AI agents can substitute journalists as interviewers and scale the number of interviewees.
[176] Bastian, M., Makhortykh, M. and Dobber. T. (2019), 'News personalization for peace: how algorithmic recommendations can impact conflict coverage', *International Journal of Conflict Management*, 30(1), DOI: 10.1108/IJCMA-02-2019-0032.
[177] Evergreen content is continually relevant, and stays 'fresh' for a much longer period of time than breaking news articles for example.

6. Sociotechnical Systems in Context

> Automation is a reality, but it is also an ideology.
> – Astra Taylor, 2018.[178]

The social and political context in which automation takes place is crucial. Researchers have warned that in contrast to other cyberthreats, the vulnerabilities of AI and ML are not merely touchpoints (online passwords, keys, etc.); they also exist 'in the interactions within and between the social, cultural, political, and technical elements of a system'.[179] To overcome the paradox of quantifying the qualitative – discourses or social interactions – it is critical to address AI and ML as sociotechnical systems.[180] Similarly, the common fixation with 'embodied' versions of AI technologies (robots, etc.) tends to serve as a distraction from less tangible manifestations, such as algorithms and adaptive complex systems.[181] AI technology's invisibility often encourages individuals either to discount its real-world implications or, conversely, to exaggerate its risks.

In an era in which all aspects of political and social life will be transformed by AI, the current multilateral and interactive channels of communication needs to be examined not just as discourse but also 'as data collection, storage and processing'.[182] In that context, personalization has emerged as a model of communication that aligns with the restructuring of digital advertising so as to prioritize quantifiable online engagement. Like every technology, personalization comes with embedded goals, and it is incumbent on the agents employing it to evaluate whether their own goals are compatible,[183] or whether they can achieve AI alignment.[184] AI-driven transformations call for a meaningful, contextually aware and culture-specific debate about human values and the social contract that protects them, and a re-evaluation of delicate balances that sustain democratic societies, such as that between the public and private sphere.

Where ethics comes in

AI may not come to dominate humankind, but it will penetrate and transform every aspect of our lives, from public services, to personal communication, to employment. But the data on which AI relies are abstractions of real-life phenomena, not an objective representation of the world, and they carry their own bias. Supervised ML algorithms, by prioritizing certain functions over others while being trained, can exhibit their own learning bias. In that context, and given the scale of the events

[178] Taylor, A. (2018), 'Talk: Fauxtomation', AI Now 2018 Symposium, New York, 16 October 2018, https://ainowinstitute.org/symposia/videos/fauxtomation.html (accessed 4 Jul. 2019).
[179] Elish, M. C. and Watkins, E. A. (2019), 'When Humans Attack: Re-thinking safety, security, and AI', Medium, 14 May 2019, https://points.datasociety.net/when-humans-attack-re-thinking-safety-security-and-ai-b7a15506a115 (accessed 24 May 2019).
[180] In July 2019 the Knight Foundation announced a large grant in support of research aiming to establish a new field investigating the impact of technology on democracy. See https://knightfoundation.org/press/releases/knight-fifty-million-develop-new-research-technology-impact-democracy (accessed 23 Jul. 2019).
[181] Avin, S. (2019), 'Exploring artificial intelligence futures', *Journal of Artificial Intelligence Humanities*, 2, https://www.shaharavin.com/publication/pdf/exploring-artificial-intelligence-futures.pdf (accessed 6 Jul. 2019).
[182] Langlois, G. and Elmer, G. (2013), 'The research politics of social media platforms', *Culture Machine*, 14: p. 2.
[183] Helberger, N., Bodó, B., Sørensen, J. K. and van Drunen, M. Z. (2018), *News personalization symposium report*, University of Amsterdam, Institute for Information Law, 5 May 2018, http://personalised-communication.net/wp-content/uploads/2018/05/Report-2018-Amsterdam-News-Personalisation-Symposium-1.pdf (accessed 6 Jul. 2019).
[184] More specifically, AI alignment is defined as 'the task of ensuring that artificial intelligence systems reliably do what humans want'. See Irving, G. and Askell, A. (2019), 'AI Safety Needs Social Scientists', OpenAI, 19 February 2019, https://distill.pub/2019/safety-needs-social-scientists/#learning-values-by-asking-questions (accessed 27 May 2019).

and processes these systems will be affecting, it is essential to heed warnings that social inequalities or patterns of oppression that algorithmic systems are in danger of not merely perpetuating, but exacerbating, will need to be addressed urgently and comprehensively.[185] Endemic biases and patterns of discrimination across various domains, from media to work environments, can become even more entrenched when, encoded in algorithmic systems, they become abstract and obfuscated.

Algorithms are value-laden, and the scope of their cross-border and intersectoral employment calls for the incorporation of human-centric values. Nevertheless, concerns have been raised about the risks of fixating on the 'universality' of ethical frameworks that may not just be impossible to achieve, but may also detract from a meaningful, contextually aware and domain-specific investigation of algorithmic systems that could challenge certain business interests.[186] Scale is not always the most efficient way to optimize, especially in relation to systems that can impact human rights and the resilience of political systems. Domain knowledge is vital, as it will dictate, among other things, a model's level of abstraction or appropriate oversight mechanisms. In that context, the HLEG on Artificial Intelligence's *Ethics guidelines for trustworthy AI*, commissioned by the European Commission, have raised some vital issues, such as the need to adopt a sectorial approach, to make sure AI systems are auditable and human-centric, and to address asymmetries of power or information to which these systems may give rise.[187]

Nevertheless, ethical guidelines are not a cure-all for malicious uses or the unintended consequences of AI, and as the list of ethical guidelines expands, it is important to understand that this 'hyperactivity' belies other risks.[188] Regrettably, AI ethics is an extremely malleable concept[189] that, in the absence of meaningful oversight and accountability mechanisms, lacks currency.

A call to employ the normative power of journalistic codes of ethics

Bearing in mind the limitations of ethics in respect to the application of AI, professional norms – such as journalistic codes of ethics – can play 'a productive role in concert with other AI governance schemes, including legal requirements and safeguards'.[190] That is why this paper calls not just for an overhaul of editorial codes but also for a broader review of the information space in which personalization systems attempt to embed themselves. Algorithmically-driven automation is liable to exacerbate power asymmetries not only between audience and media conglomerates, but also between journalists and the commercial imperatives of the companies they work for. Journalistic

[185] Sloane, M. (2019), 'Inequality Is the Name of the Game: Thoughts on the Emerging Field of Technology, Ethics and Social Justice', Proceedings of the Weizenbaum Conference 2019 'Challenges of Digital Inequality – Digital Education, Digital Work, Digital Life', Berlin, doi: 10.34669/wi.cp/2.9, pp. 1–9 (accessed 11 Sept. 2019).
[186] Veale (2019), 'Governing Machine Learning that Matters', p. 54.
[187] European Commission: Independent High-Level Expert Group on Artificial Intelligence (2019), *Ethics guidelines for trustworthy AI*, European Commission, DG Connect, 8 April 2019, https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (accessed 3 Jun. 2019).
[188] Risks include the malpractice of actors being able to cherry-pick the framework best suited to retrofitting their pre-existing behaviours, using guidelines to implement superficial changes or to delay legislation, and others. See Floridi, L. (2019), 'Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical', *Philosophy & Technology*, 32(2): pp. 1–9, https://link.springer.com/article/10.1007%2Fs13347-019-00354-x (accessed 27 May 2019).
[189] Calo, R. (2017), 'Artificial Intelligence Policy: a Primer and Roadmap', SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3015350, p. 7 (accessed 2 Jun. 2019).
[190] Gasser, U. and Schmitt, C. (2019), 'The Role of Professional Norms in the Governance of Artificial Intelligence', to appear in: Dubber, M. D., Pasquale, F. and Das, S. (eds.) (forthcoming), *The Oxford Handbook of Ethics of AI*, Oxford University Press, http://dx.doi.org/10.2139/ssrn.3378267 (accessed 30 Jun. 2019).

ethics, as context- and sector-sensitive codes, can steer human- and citizen-centric AI deployment in digital media, provide direction for technology companies, and inform and expand oversight and accountability mechanisms already embedded in national regulatory frameworks.

AI ethics, as they relate to news and journalism, should mix 'institutionalized codes, professional cultures, technological capabilities, social practices, and individual decision making'.[191] Journalists and editorial teams around the world should urgently reassess and rearticulate their ethical guidelines and codes of practice, to ensure that they are fit for purpose, and set a robust framework of digital transition for legacy media that balances innovation with core journalistic principles such as impartiality, inclusiveness, diversity, objectivity, fair reporting and the public interest. Those principles will need to be refined and re-evaluated in a way that makes them relevant to the information age, and that sets a clear direction for the ML models that will seek to encode them.

Legacy media's norms and ethics that relate to the public interest or inclusiveness may not be present in new platforms and aggregators[192] – the 'liminal press'[193] – that tend to prioritize engagement and consumer needs. For example, Ananny and Crawford's interviews with 'liminal press' designers indicated that app design and news values are not really intersecting.[194]

Current editorial codes can be used as a basis to incorporate AI ethics. A 2015 study into national journalistic codes of ethics across the world found that 91 per cent of the codes surveyed lacked references to the digital environment;[195] it is clear that legacy media have a lot of ground to cover. In the UK, the BBC is governed by the corporation's Royal Charter,[196] which defines its mission as (among other things) the provision of 'duly accurate and impartial news' to 'build people's understanding of all parts of the United Kingdom and of the wider world', to 'raise awareness of the different cultures and alternative viewpoints that make up its society', and to 'help contribute to the social cohesion and wellbeing of the United Kingdom'.

The BBC's editorial guidelines also dictate that it has to act in the public interest and avoid misleading audiences.[197] Its public service role sits at the centre of its adoption of ML – and more specifically, of personalization – and debates as to the purpose for which its recommender systems will optimize are ongoing, but according to one senior executive within the corporation, the aim is to provide services that encode serendipity and human co-curation,[198] that encourage rich experiences and that are editorially responsible. The BBC Datalab and technology forecasting teams are working on ML guidelines intended to inform the corporation's main editorial guidelines, and

[191] Ananny, M. (2016), 'Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness', *Science, Technology & Human Values*, 41(1): p. 96.
[192] Diakopoulos, N. and Koliska, M. (2016), 'Algorithmic Transparency in the News Media', *Digital Journalism*, doi: 10.1080/21670811.2016.1208053 (accessed 2 Jun. 2019).
[193] Ananny, M. and Crawford, K. (2015), 'A liminal press', *Digital Journalism*, 3(2): pp. 192–208, doi: 10.1080/21670811.2014.922322 (accessed 2 Jun. 2019).
[194] Ibid.
[195] Bosnia and Herzegovina, Canada, Hungary, Luxembourg, the Netherlands, Norway, Poland, Romania, and the UK are exceptions. See Díaz-Campo, J. and Segado-Boj, F. (2015), 'Journalism ethics in a digital environment: How journalistic codes of ethics have been adapted to the Internet and ICTs in countries around the world', *Telematics and Informatics*, 32 (4): pp.735–44, doi: 10.1016/j.tele.2015.03.004 (accessed 11 Sep, 2019).
[196] BBC (2019), 'About the BBC: Charter and Agreement', https://www.bbc.com/aboutthebbc/governance/charter (accessed 17 Nov. 2019).
[197] BBC (2019), *Editorial Guidelines*, http://downloads.bbc.co.uk/guidelines/editorialguidelines/pdfs/bbc-editorial-guidelines-whole-document.pdf (accessed 11 Sept. 2019).
[198] Author's email exchange with Laura Ellis, head of technology forecasting at the BBC, 18 July 2019.

have already released a set of principles for the corporation's use of ML.[199] The *Guardian*'s editorial guidelines underline the importance of preventing the public from being misled and distinguishing between comment, conjecture and fact.[200] While social media have complicated this task – common to many editorial codes, including that of the UK's Independent Press Standards Organisation (IPSO)[201] – it should be factored into the development of personalization systems.

The *New York Times*'s editorial standards highlight neutrality and the need to treat readers 'no less fairly in private than in public'.[202] The question which then arises is this: is personalized recommendation a public or a private form of communication, and what constitutes fair treatment in that context? The different ways to measure fairness outlined by Veale[203] can help orient the debate, but cannot be extrapolated to the media environment before the debate has actually taken place about what fairness means in an ML-augmented media ecosystem.

The *Washington Post* affirms its commitment to fairness, a concept that it states includes completeness (no important facts are omitted), relevance and honesty.[204] As already noted, unpacking what journalistic principles mean in practice can both help instil more transparency in a personalization system and help model it. It is also important to regard the concept of completeness as a lens on understanding long-term news events. After all, most politically, socially and environmentally impactful events do not simply happen overnight. In that context, recommender systems should include content that provides a complete picture of an event or a debate.

[199] Straub, G. (2019), 'Scaling responsible machine learning at the BBC ', BBC, 4 October 2019, https://www.bbc.co.uk/blogs/internet/entries/4a31d36d-fd0c-4401-b464-d249376aafd1 (accessed 14 Oct. 2019).
[200] *Guardian* (2011), *Editorial Guidelines: Guardian News & Media Editorial Code*, August 2011, https://uploads.guim.co.uk/2018/10/20/273521476.pdf (accessed 8 Jul. 2019).
[201] The Independent Press Standards Organisation is the current UK regulator for newspapers and magazines; it replaced the Press Complaints Commission in 2014.
[202] *New York Times* (2019), 'Ethical Journalism: A Handbook of Values and Practices for the News and Editorial Departments', https://www.nytimes.com/editorial-standards/ethical-journalism.html (accessed 8 Jul. 2019).
[203] Veale mentions statistical/demographic parity, accuracy equity, conditional accuracy equity, equality of opportunity, and disparate mistreatment. See Veale (2019), 'Governing Machine Learning that Matters', pp. 68–69.
[204] *Washington Post* (2016) 'Policies and Standards', 1 January 2016, https://www.washingtonpost.com/news/ask-the-post/wp/2016/01/01/policies-and-standards (accessed 8 Jul. 2019).

7. Safeguarding a Free Press and Algorithmic Sorting in a Speculative Future

> Open societies need to regulate companies that produce instruments of control, while authoritarian regimes can declare them "national champions."
> – Soros, G., 2019.[205]

Recent sociopolitical trends, such as disinformation or the escalating polarization of the digital space, have illustrated why it is critical to move beyond the idea that regulating cyberspace is difficult. In a sense, regulation is already taking place. Digital platforms' architectures, and the AI systems transforming them, can be seen as attempts to regulate in an indirect way[206] that – until recently – shielded them from suffering political cost. Regulating AI at such an early stage is not an easy task,[207] but as a baseline, a deep understanding of the fundamentals of AI and the issues to which it can give rise is necessary for policymakers. Developing the expertise to study its social and political impact, even on a speculative basis, is fundamental.

AI is challenging, not just because it poses novel problems, but because it also demands novel approaches. Transparency, for example, even though often cited as a safeguard against abuses, and highlighted as an imperative by reports on AI,[208] is not, in itself, an adequate way to govern algorithmic systems. Researchers indicate policymakers need to look not just *inside* systems, but *across* them,[209] in order to properly diagnose the relationship between human and non-human actors rather than just an internal logic. The global financial crisis that began in 2007 is an example of the inability to do exactly that. Any form of transparency has always to be paired with accountability and oversight, and carefully evaluated, as the flipside to a 'digitally afforded public transparency' can be 'digitally enabled surveillance'.[210]

Human rights law is becoming the focus of technology regulation debate, as a universally binding set of standards that is 'well-suited for borderless technologies'.[211] In a way, it makes sense that the unprecedented, truly global reach of digital intermediaries might provide a real test for the provisions of international human rights treaties. In a time when global governance systems are challenged by recent political shifts, tackling rising sociotechnical issues such as the adoption of AI-driven personalization becomes even more challenging. Ultimately, personalization may impinge on our ability to address digitally-produced harms that manifest at group level. A report from the

[205] Soros, G. (2019), 'Remarks delivered at the World Economic Forum', 24 January 2019, https://www.georgesoros.com/2019/01/24/remarks-delivered-at-the-world-economic-forum-2 (accessed 28 Jun. 2019).
[206] See Lessig's modalities of regulation: law, norms, market and architecture, in Lessig, L. (1998), 'The New Chicago School', *The Journal of Legal Studies*, The University of Chicago Press, 27(S2): pp. 661–91.
[207] For a suggested roadmap see Calo (2017), 'Artificial Intelligence Policy: a Primer and Roadmap'.
[208] Including the conference report of *Governing the Game Changer – Impacts of artificial intelligence development on human rights, democracy and the rule of law*, conference organized by the Council of Europe in Helsinki, 26–27 February 2019, https://rm.coe.int/conference-report-28march-final-1-/168093bc52 (accessed 6 Jul. 2019).
[209] Ananny, M. and Crawford, K. (2018), 'Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability', *New Media & Society*, 20(3): pp. 973–89, doi: 10.1177/1461444816676645 (accessed 11 Sept. 2019).
[210] Anderson, C. W. (2012), 'Towards a sociology of computational and algorithmic journalism', *New Media & Society*, 15(7): doi: 10.1177/1461444812465137 (accessed 11 Sept. 2019), p. 1011.
[211] Amnesty International and Access Now (2018), 'The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems', Access Now, 16 May 2018, https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/ (accessed 2 Dec. 2019).

Council of Europe quite astutely called for a reconsideration of our current conceptions of human rights and the mechanisms through which they are enforced.[212]

Looking at AI adoption in digital communication, a 'bottom-up' sectorial regulatory approach is needed, coordinated to avoid incompatibility in rules from different sectors, and complemented by broader, national-level regulation. The barrister and writer Jacob Turner has suggested the idea of a pyramid structure, with 'high-level standards at the top and then an increasing number of lower-level bodies', and innovative approaches to regulation, such as the UK Financial Conduct Authority's regulatory 'sandbox'.[213] The idea of sandboxes was also put forward by the European Commission-appointed HLEG.[214] Jurisdiction is one of the key issues in regulating AI systems, but putting in place a network of measured but robust regulatory oversight that cascades from the sector-specific to the domestic and to the international level would go a long way towards addressing this issue.

As a general point, personalization both in legacy and social media should be clearly stated, and citizens should be able to move between different layers of personalization.[215] Wang and Diakopoulos have also suggested – as a potential safeguard against filter bubbles – the designing of content personalization platforms that focus on clusters rather than individuals,[216] but even that direction should take into consideration group privacy and discrimination implications. Individuals should have easy access to their data when those are collected, stored and processed for the purposes of building and deploying personalized algorithmic systems. In terms of targeted advertising, the trade-offs between individual and contextual targeting – the latter focusing on the surrounding content a user is exposed to, rather than on the user themselves – should be examined and considered under current and incoming data protection and privacy regulation.

The following high-level recommendations for governments, legacy media (news outlets, digital natives, broadcasters, etc.), algorithmic systems, their engineers, and digital intermediaries, are forward-looking, seeking to prevent negative impacts on societies' foundations and the individual. Nevertheless, both historical and prospective responsibility should be addressed.[217]

[212] Yeung, K. (2018), A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework, Council of Europe, Committee of Experts on Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence (MSI-AUT), 9 November 2018, p. 70.
[213] Interview with the author, London, 18 March 2019.
[214] European Commission: Independent High-Level Expert Group on Artificial Intelligence (2019), *Policy and investment recommendations for trustworthy Artificial Intelligence*. For a critique of the report's shortcomings including the absence of challenges to corporate governance, see Veale, M. (2019), 'A critical take on the policy recommendations of the EU High-Level Expert Group on Artificial Intelligence', European Journal of Risk Regulation (forthcoming).
[215] Diakopoulos, N., Cass, S. and Romero, J. (2014), 'Data-driven rankings: the design and development of the IEEE top programming languages news app', Proceedings of Symposium on Computation + Journalism.
[216] Wang, Y. and Diakopoulos, N. (2018), 'Considerations for Article-Level Personalization of News Content', Algorithmic Personalization and News (APEN18): Workshop at the 12th International AAAI Conference on Web and Social Media, Stanford, CA, 25 June 2018.
[217] Yeung, K. (2018), p. 46.

Recommendations for governments

Existing regulatory frameworks, standards and laws that pertain to AI adoption in the information space need to be evaluated, but extrapolating them may not be sufficient for the purposes of establishing meaningful and effective oversight mechanisms. The efficiency of regulatory and legal frameworks needs to be reviewed systematically to ensure they are fit for purpose for constantly evolving ML systems and their potential risks. As first steps:

- Current national media regulators, such as the Office of Communications (Ofcom) in the UK or the Federal Communications Commission (FCC) in the US, should expand their oversight to algorithmic systems used in personalization by each creating a new unit specializing in algorithmically-curated or -produced communication. Redress mechanisms for consumers should be put in place, and proportionate sanctioning power should be vested in the regulator. A fit-for-purpose policy demands coordination with national election commissions and data protection authorities to protect citizens' privacy and the integrity of democratic processes, as the take-up of personalized communication by political actors is accelerating.

- Carefully consider what models of responsibility and legal liability are appropriate for media actors of varying scale, reach and ownership, employing different forms of personalization. If the 'duty of care' is to be adopted,[218] it has to be refined to fit the current multinodal, multi-layered and interactive information environment. Specific compliance and redress timelines following rights violations also need to be instituted.

- Earmark funding for human–computer interaction (HCI) research into how citizens interact with news, under the auspices of the new unit of the media regulator. The establishment of a robust evidence base for the societal, psychological and political effects of personalization systems in communication should be a priority, so that targeted policy interventions can have practical and long-term effects. Governments should, moreover, consider making running pre-release trials of algorithmic systems mandatory, with clear benchmarking.

- Allocate funding to research into current accountability mechanisms between platforms, legacy media, AI and ML developers and data brokers, as they relate to data and ML model transfers and augmentation.

- Despite the positive steps by certain tech firms in relation to political advertising, online advertising demands close scrutiny by regulators, encompassing adtech, social media companies, legacy media. Governments should also review advertising regulation *vis-à-vis* targeting, and audit who is targeted, bearing in mind both individual and group rights.

[218] The UK government's Online Harms White Paper, for example, proposed establishing in law a duty of care towards users, which would be overseen by a new regulator. This is a concept which is also gaining ground within the EU Commission.

- Promulgate the importance of data infrastructure literacy[219] that is predicated on transparency in terms of data collection, inference production and audience segmentation, as well as processes for data subjects to exercise their rights.
- Review media regulation to ensure media plurality is sustained across personalized systems.
- Explore standardization measures for personalized systems such as recommender systems and ad targeting, and consider making AI engineering a regulated profession.
- Engage national media regulators with counterparts on the issue of AI in media under the auspices of the International Telecommunication Union (ITU), during the ITU Global Symposium for Regulators (GSR). And regulators should engage with the Institute of Electrical and Electronics Engineers (IEEE) Standard Association's AI standards series, and establish channels of communication on AI development and deployment with national standardization agencies.

Recommendations for legacy media and digital-native publishers

Legacy media and digital-native publishers dedicated to the protection of the public interest need to defend their legitimacy. An effective uptake of ML personalization in the newsroom would entail the following action points:

- Refine what the term 'public interest' means, and how it can be best served in an AI world, before articulating the targets of ML algorithms employed in personalization.
- Consider how current journalistic ethics such as inclusiveness, impartiality and fair reporting can be reflected in the building and deployment of ML systems that guarantee content diversity as well as balanced and evidence-based analysis. AI and ML ethical guidelines should be embedded in publicly accessible ethical codes and editorial guidelines, disclosing any potential trade-offs that have been made in model development. The reviewed guidelines should incorporate provisions for algorithmic personalization in the context of elections and democratic processes.
- Position transparency at the centre of personalization tools, by allowing users to modify their personalization within constraints predetermined by risk analysis and data protection impact assessments, or provide an opt-out option.[220] Users should be made aware of what the algorithmic systems to which they are exposed are optimizing for.
- Adopt clear and easy-to-read policies disclosing the use – if any – of third-party data and digital signal collection in training ML models.

[219] Gray, J., Gerlitz, C. and Bounegru, L. (2018), 'Data infrastructure literacy', *Big Data & Society*, July–December 2018, pp. 1–13, doi: 10.1177/2053951718786316 (accessed 11 Sept. 2019).
[220] Wang and Diakopoulos (2018).

- Senior management and editors will need to acquire baseline AI knowledge as an important step in addressing the asymmetry of information with data science teams, the information arbitrage between them and social media companies, and to improve both their editorial decisions and strategic planning.

- Refine ML pipelines to ensure the monitoring and maintenance of systems, as well as lines of accountability, to avoid the trap of 'many hands'.[221]

- Prioritize addressing diversity issues in the data science and editorial recruitment processes.

Auditing algorithmic systems

As noted earlier, bearing in mind that algorithms can be transient, being transformed through their interaction with other algorithms, users, data, etc.,[222] policymakers need to establish not just a regulatory framework of risk management, but also one that is regularly reviewed and revised. The following points should be considered within any oversight of algorithmic systems used in personalization:

- Following the recommendations of the UN Secretary-General's High-level Panel on Digital Cooperation, audit and certification schemes to monitor compliance of AI systems with engineering and ethical standards should be considered in the media environment.[223] AI systems should be audited for their learning and data biases.

- Checks should be built into the system, monitoring when personalization systems need retraining, adjusting or rebuilding to overcome concept drift.

- Human rights impact assessments drawing on international human rights, data protection and privacy rights should be employed.

- This paper agrees with AI Now's recommendation for a 'full stack supply chain' accountability and transparency mechanism that follows the entire ML development pipeline – logging the source of the training data, the inferred attributes pursued, the models used and any AI application programme interfaces (APIs) used – that could be screened by the appropriate regulator.[224] Apart from the research field of data provenance, the concept of decision provenance[225] also merits further attention. The 2019 report published by the Alan Turing

[221] Technology philosopher Helen Nissenbaum used this term to denote the difficulty of attributing blame in complex systems involving many actors, organizations and components.
[222] Harambam, J., Helberger, N. and van Hoboken, J. (2018), 'Democratizing algorithmic news recommenders: how to materialize voice in a technologically saturated media ecosystem', 376, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, p. 4, doi: 10.1098/rsta.2018.0088 (accessed 11 Sept. 2019).
[223] UN Secretary-General's High-level Panel on Digital Cooperation (2019), *The Age of Digital Interdependence*, p. 5, https://www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf (accessed 11 Sept. 2019).
[224] Richardson, R. (2019), 'Optimizing for Engagement: Understanding the Use of Persuasive Technology on Internet Platforms', written testimony to the US Senate Committee on Commerce, Science and Transportation: Subcommittee on Communication, Technology, Innovation and the Internet, 25 June 2019, p. 8, https://ainowinstitute.org/062519-richardson-senate-testimony.pdf (accessed 30 Jun. 2019).
[225] The concept introduced by Cobbe, Singh and Norval as a means to improve built-in accountability of algorithmic systems relates to information on decision pipelines. See Cobbe, J. Singh, J. and Norval. C. (2018) 'Decision provenance: harnessing data flow for accountable systems', IEEE Access, Vol. 7, pp. 6562–6574, doi: 10.1109/access.2018.2887201.

Institute on responsible design and implementation of AI systems for the public sector could be drawn on in setting up the broader governance framework.[226]

- Repurposing of algorithmic models should be strictly controlled, so that personalization systems built on one company's audience data are not augmented with systems addressing different audiences. The trend towards tradeable, augmentable algorithmic models, or 'learnware',[227] demands urgent attention.

- Algorithmic systems in communication should enhance users' agency over them by clearly identifying themselves as such, making it clear that humans are not controlling the curation of content. User feedback mechanisms should also be incorporated.

Training, recruiting and overseeing the AI engineers of the future

AI engineers play a central role in designing the algorithmic systems, in terms of deciding what data attributes to include, how to weigh them, what models to use and how to employ feature engineering.[228] To some extent, the bias that drives individual decisions can be counterbalanced by diversity in the recruitment process. As the AI Now Institute has warned in a 2017 report, bias can be encoded in AI systems because of a lack of diversity in the group of individuals creating them,[229] with an average of 80 per cent of AI professors at leading computer science universities currently being male.[230] The UN has also called for gender equality and inclusion of marginalized groups,[231] while AI Now has emphasized that discrimination relates not only to sex, but also to gender, sexual orientation and race.[232] For these reasons:

- Ethics should be established as a mandatory component of any data science course.

- Certification schemes should be considered. Training of AI engineers should scale and diversify in relation to the skill they would certify for. Builders of systems that will have global reach should have a broader knowledge base that encompasses social sciences and a curriculum commensurate with the scale of their responsibilities and impact.

- The current lack of gender, racial and social diversity within the AI community should be addressed urgently.

- The production line of AI systems used in communication and their accountability structures should be refined, with accountability by design being embraced.[233]

[226] Leslie, D. (2019), *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*, Alan Turing Institute, p. 25, https://doi.org/10.5281/zenodo.3240529 (accessed 26 Nov. 2019).
[227] Binns, R., Edwards, L. and Veale, M. (2018), 'Algorithms that remember: model inversion attacks and data protection law', 376, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, pp. 3–4, doi : 10.1098/rsta.2018.0083 (accessed 11 Sept. 2019).
[228] Feature engineering in ML is the process of transforming data inputs into new data that will, in turn, be used in the model. See Veale (2019), 'Governing Machine Learning that Matters', p. 38.
[229] AI Now Institute (2017), *AI Now 2017 Report*, pp. 16–18, https://ainowinstitute.org/AI_Now_2017_Report.pdf (accessed 6 Jul. 2019).
[230] AI Index 2018 (2018), *Artificial Intelligence Index 2018*, p. 25, https://aiindex.org.
[231] UN Secretary-General's High-level Panel on Digital Cooperation (2019), *The Age of Digital Interdependence*, p. 29.
[232] West, Whittaker and Crawford (2019), Discriminating Systems: Gender, Race and Power in AI.
[233] Leslie (2019), p. 25.

Regulating technology companies

The past years have proved without doubt that self-regulation in the tech sector has limited potential. That reality, combined with the unprecedented normative power of big tech, calls for a change of direction in the habitual *laissez-faire* approach. Underlying incentive structures need urgent evaluation as well as horizontal and vertical mergers, as they tend to impact on data sharing and augmentation schemes. Steps that should be prioritized are:

- The norm-setting regulatory power of companies with strategic market status needs to be scrutinized. Competition law should be approached as a process, and examine reform that spots anti-competitive practices that move beyond price differentials – the tech giants' services are mostly delivered 'free of charge' – to encompass data power and infrastructural capture.

- Mergers need to be properly scrutinized, bearing in mind the network effects that can transform a medium-sized firm into a scaled-up operation. Data transfers through mergers and acquisitions should be scrutinized, too.

- Trade-secret legal protections of algorithmic communication should be reassessed. Policymakers should work towards establishing a framework that can make these models visible and explainable to auditors.

- Citizens' consent to data storage and processing has to be meaningful, and not framed as a zero-sum option. The trade-off between privacy-as-confidentiality and privacy-as-control[234] has to be transparent to data subjects.

- Technology companies should not be allowed to retroactively change data privacy policies to boost their personalization systems without meaningful approval by regulators and data subjects.

- The sanctioning and penalizing of legal or regulatory breaches needs to become more innovative, involving more than monetary fines that can become normalized as 'part of doing business'. More effective measures may include the withdrawal of licences to operate, embedding of regulatory teams, and policies that severely curtail data collection.

[234] Veale has argued that de-identifying methods used by tech companies often have the side-effect of depriving data subjects of meaningful control over their data. See Veale (2019), 'Governing Machine Learning that Matters', p. 138.

8. Conclusion

AI can 'both supplement and replace human decision-making',[235] but the potential scale of its effects calls for thorough examination of its implementation. Similarly, AI-driven personalization in communication merits meaningful examination as to why, where and how it is deployed, as it can effect a value drift, transform how individuals relate to society and threaten democratic norms by changing how political campaigning is enacted. There is an urgent need to develop an ethical framework to define how the technology is deployed in the media environment, given the scope for manipulation that is able to disrupt social and political cohesion and impinge on individual and group rights. The window of opportunity is narrowing, as technology companies, adtech and e-commerce business norms are crowding out the editorial guidelines and statutory law that have traditionally defined the norms and legal framework of privacy rights, public debate and deliberation.

Personalization seems to reflect a broader trend in the information space, in that it assigns responsibility to the individual (i.e. the consumer) for being aware of disinformation through personal digital literacy, and/or for remaining cognizant of technology companies' data harvesting policies by navigating overwhelmingly complex terms and conditions. But this approach challenges the long-standing model of legacy media and their mission statement: serving the public interest. Media's institutional environments and decision-making cultures matter, because they affect how citizens view the world. The regulatory framework through which legacy media had to operate – a combination of norms and statutory regulation – provides a reliable foundation to approach the coalescing of their communication strategies with those of technology companies.

While research suggests that certain platforms may not be conducive to the Habermassian concept of a public sphere,[236] it is important to avoid deterministic views of technology. In sociotechnical systems, examining the social is as important as examining the technical. As Selbst et al. have suggested, we 'draw our analytical boxes around both human and technical components'.[237] Even if the concepts of filter bubbles and echo chambers remain contested, they have raised serious concerns about the potential ramifications of the digital ecosystem. Polarization, discrimination and marginalization are not phenomena specifically created by Web 2.0, but AI technologies can amplify and entrench them by encoding them in digital infrastructures, transforming communication processes and accepted norms and values.

Value systems are in constant negotiation through discourse, meaning they can also be influenced by shifts in power over how people receive and impart information. Shifts may not be seismic, but long-term and undetected. Personalization creates too many scalable risks to be deployed uncritically. ML system deployment should never be seen as a foregone conclusion. Rather, it

[235] Inkpen, K., Veale, M., Chancellor, S., De Choudhury, M. and Baumer, E. P. S. (2019), 'Where Is the human? Bridging the gap between AI and HCI',CHI '19, Extended Abstracts, 4–9 May 2019, Glasgow, p. 2.
[236] Furman, I. and Tunç, A. (2019), 'The End of the Habermassian Ideal? Political Communication on Twitter During the 2017 Turkish Constitutional Referendum', *Policy & Internet*, doi: 10.1002/poi3.218 (accessed 11 Sept. 2019).
[237] Selbst, A. D., boyd, d., Friedler, S. A., Venkatasubramanian, S. and Vertesi, J. (2018), 'Fairness and Abstraction in Sociotechnical Systems'.

should be adjusted to take account of the particular context and time-specific trade-offs. At times, the implication will be 'not to design', or not to deploy.[238]

AI-driven personalization may not be a 'clear and present danger', but it is a sociotechnical system that policymakers and media professionals need to get right, and its application will need to be approved on a case-by-case basis. The alternative – i.e. uncritical adoption – could bring states face to face with unforeseen social and political crises rooted in highly opaque communication networks that reinforce biases, obscure discrimination, and undermine the transparency that underpins democracy.

[238] Baumer, E. P. S. and Silberman, M. S. (2011), 'When the implication is not to design (technology)', CHI '11, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2271–74.

About the Author

Sophia Ignatidou joined Chatham House in 2018 as an Academy Stavros Niarchos Foundation Fellow in the International Security Department. She researches artificial intelligence, disinformation, political campaigning, propaganda and surveillance. She previously worked as a freelance journalist and digital sub-editor for the *Guardian*, the *Sunday Times* and CNN, among others. Sophia holds an MA in journalism from Goldsmiths, University of London, and an MA/PGDip in international studies and diplomacy from the School of Oriental and African Studies, University of London.

Acknowledgments

The author would like to thank Shahar Avin, Reuben Binns, Kourtney Bitterly, Ryan Budish, Corinne Cath, Jennifer Cobbe, Laura Ellis, Stephen Fozard, Dan Gilbert, Ana Jakimovska, Jonas Kaiser, John Keefe, Jaakko Lempinen, Mia Shuang Li, Francesco Marconi, Paul Nemitz, Lisa-Maria Neudert, Nic Newman, Bruno Freitas de Oliveira, Titus Plattner, Mona Sloane, Jacob Turner and Karina Vold for comments and discussions on which this work is based. Equally valuable were talks and panel discussions at CPDP 2019 in Brussels, Data Science Salon New York, and workshops at the Alan Turing Institute, London.

Thanks are also due to Calum Inverarity, Chronis Kapalidis, James Kearney, Patricia Lewis, Beyza Unal, Janet Waters, and the 2019 Chatham House Academy team and fellows for their valuable support and help.

Independent thinking since 1920

Cover image: The Reuters and other news apps seen on an iPhone, 29 January 2019.

Photo credit: Copyright © NurPhoto/Contributor/Getty

ISBN 978 1 78413 373 3

This publication is printed on FSC-certified paper.

FSC

Typeset by Soapbox, www.soapbox.co.uk

The Royal Institute of International Affairs
Chatham House
10 St James's Square, London SW1Y 4LE
T +44 (0)20 7957 5700 F +44 (0)20 7957 5710
contact@chathamhouse.org www.chathamhouse.org

Charity Registration Number: 208223