The changing face of Russia's information war against Ukraine and other democratic countries: Lessons and recommendations. Interview with Professor Sinan Aral

Sinan Aral

https://orcid.org/0000-0002-2762-058X

Massachusetts Institute of Technology, MIT Sloan School of Management, 100 Main St., Cambridge, MA 02142, USA

Domalewska: You have carried out extensive research on misinformation and disinformation that has helped us understand how fake news is distributed and what strategies are implemented to distribute it on social media. You have proved that "[f]alsehood [was] diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects were more pronounced for false political news than for false news about terrorism, natural disasters, science, urban legends, or financial information" (Vosoughi *et al.*, 2018, p. 1146). This means that false political news was more viral, spread deeper and diffused more broadly than any other category of fake news (Aral, 2020, p. 47). Social bots distributed both true and false information at the same time, which suggests that it is humans who are responsible for spreading fake news rather than bots and there are several reasons for that. First, fake news is more novel than the truth, hence it is more likely to be shared by users who want to be seen as being in the know or having access to inside information, which confers their social status (which you call the novelty hypothesis). Second, by examining users' replies to

information, you found that fake news evoked fear, disgust, and surprise whereas true stories inspired anticipation, sadness, joy, and trust (Vosoughi *et al.*, 2018, p. 1150). Emotional content has more economic value as it generates greater user engagement. Posts that evoke emotional arousal, both positive (such as awe) and negative (such as anger or anxiety) are more viral than posts that evoke low-arousal or deactivating emotions (such as sadness) (Berger and Milkman, 2012). Engaging content is then amplified by algorithms whose role is to boost the popularity of captivating posts in order to engage users. In this way, a 'tyranny of trends' is created (Aral, 2020, p. 215). The 2016 Presidential elections have shed light on Russia's involvement in democratic processes in the United States, and in other countries too. To protect our democracy, many initiatives have been launched - both legislative (e.g. the Honest Ads Act and the Secure America from Russian Interference Act of 2018 passed in the US, the Network Enforcement Act[1] in Germany, and the Law on the fight against information disorder[2] in France) and self-regulatory (e.g., initiated by the European Commission's Code of Practice on Disinformation). Has the problem of spreading falsehood been solved? Has the way in which disinformation is spread changed?

Aral: Pre-2020 strategies of misinformation and disinformation have evolved over the last 5 years but they have not been any less intense. In fact, they have potentially been at least as intense and sometimes more intense. The tactics have changed. Before 2020, the western world was not as cognisant or prepared to deal with Russian dis- and misinformation online. And so, Russian tactics were much more overt. In particular, what they would do was use bots and human trolls and sock pockets to spread mis- and disinformation. And bots were a big part of it.[3] I think that now what you see more of is the amplification of real voices because those are harder to take down. So instead of putting a false narrative out there, they would do a search for real people creating the narrative they want to promote and then amplify that narrative through reshares, retweets, likes, and amplifications as best as they can. And when the platforms look at that, they say 'we can't take this down because the originator of the post is a real human being that we can verify as real, not a Russian spy etc.' And so it's a more roundabout way of encouraging the narratives that they want to encourage rather than explicitly creating the misinformation itself and they are finding it and encouraging it and amplifying it more. I also think that the Internet Research Agency sort of went back into the shadows and you have other arms or elements of Russian online state-sponsored misinformation and disinformation that are more clandestine, more difficult to connect directly to Russia but which are still engaged in just as intense misinformation and disinformation as before. I think it's focused a lot on Eastern Europe now, not just Ukraine where it's taking place regularly, but the surrounding countries in which the battle for the expansion of NATO is happening. These countries are big targets for misinformation and disinformation because they have become the battleground for soft and hard power in Russia's fight to limit NATO's expansion. At the beginning of the war in Ukraine, it was about support for Ukraine. It was about targeting misinformation and disinformation in countries that were providing material support to Ukraine in the form of money and weapons and similar. Supporting migrants in neighbouring countries and Poland is a great example. But now, it is at least as much

---

[1] *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz – NetzDG).*
[2] *LOI n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information.*
[3] *"In 2014, Putin engaged in a more sophisticated two-pronged information war in Crimea, using Facebook and Twitter to orchestrate the flow of information online. The first prong suppressed pro-Ukrainian voices by overwhelming their social media posts with hundreds of automated fraud and abuse reports, claiming that the posts contained porn or hate speech. In response, Facebook took down the 'offending' messages and banned their authors, effectively banishing pro-Ukrainian voices from the platform. The second prong created and disseminated misinformation through false tweets, blog posts and news stories, propagating absurd tropes describing Ukrainians as 'Nazis' who were oppressing Jews and native Russian speakers" (Aral, 2022).*

focused on NATO expansion and sort of stemming the tide of public opinion on whether NATO should or should not expand. And it is much more clandestine, much less overt, much more trying to get around the platforms' algorithms and policies[4] by supporting the real voices rather than fabricating identities and trying to spread those narratives. But it isn't less intense.

Domalewska: So Russia nowadays uses a more sophisticated way of spreading disinformation and has abandoned using automatic accounts and trolls preferring to rely on human users, ordinary people who diffuse false or manipulated information in cascades.[5]

Aral: Yes, basically Russia wants to spread disinformation or misinformation as widely as possible. So instead of creating a narrative with their false accounts and spreading this narrative, they are searching for narratives that they want out there in the world that are put out by real people who are not Russian, not bots, and they are trying to amplify those narratives and spread these narratives in cascades.

Domalewska: Disinformation that is spread and amplified by social media poses a substantial threat to democracy. Combined with declining trust in the media and politicians, it further leads to democratic backsliding.[6] What factors, in your opinion, pose a threat to democracy?

Aral: I think there are 3 prongs or 3 trends or forces that are intertwining very tightly to create a perfect storm that is threatening democracy. One is the mis- and disinformation that we were talking about. The other one is polarisation, which is on the rise in multiple different countries when the right and left are expressing more extreme views and finding less and less common ground. And then part and parcel of that is a lack of trust in the media because the media's becoming polarised. You have media that is geared towards extreme audiences on the left and on the right. They are tailoring their content for the extreme audiences and extreme opinions and that is contributing as well. So all of this is very difficult for democracy, in part because what you are seeing is a divergence of views of realities. So we don't have a common perception of the world like we did 20 years ago. All of these things together: the social media platforms recommending content in ways that sends us into divergent information spaces on one side or other of the political spectrum; the media essentially catering to these diverging audiences and opinions; political polarisation when the two sides don't listen to each other and there's a lot of animosity and hatred and something called affective political polarisation, which is rising dislike, distrust and animosity among political parties. Data on whether political polarisation is on the rise is mixed (Boxell *et al.*, 2017) whereas the data that shows affective political

[4] *Today, social media is "proactively engaged in the information war. During the Crimean annexation, they were reactive and struggled to keep up with misinformation and false abuse reports. Today, in Ukraine, they have banned Russian state-owned media from advertising on their platforms and defiantly fact-checked Putin's propaganda despite Russia's protests and a full ban of Twitter and a partial ban of Facebook in Russia. Facebook has spun up a special operations center, staffed with native Russian and Ukrainian speakers, to monitor misinformation posted about the war, added warning labels to war-related images that its software detects are more than a year old, and restricted access to content from the state-affiliated Russian media outlets RT and Sputnik. YouTube is restricting access to Russian state-owned media outlets for users in Ukraine, removing Russian state-owned channels from recommendations, and limiting their content's reach across the platform. Twitter has temporarily banned all ads in Ukraine and Russia, added labels to tweets with links to Russian state-affiliated media and downranked their content in algorithmic timelines"* (Aral, 2022).

[5] *False and manipulated content is further retweeted with each retweet spreading the news away from the original tweet, thus creating a cascade of reshares* (Aral, 2020, p. 47).

[6] *A large body of research highlights declining trust in the media and policy-makers (Collis* et al.*, 2022; Golding* et al.*, 2012), which results from widespread disinformation and fake news in the post-truth era (Kim and Gil de Zúñiga, 2020). Declining trust in the media and policy-makers combined with the populist style of governance and authoritarian values is undermining many democratic norms and practices (Norris, 2017).*

polarisation - hatred towards the other side - is on the rise is quite clear (Iyengar *et al.*, 2019). And that combined with mis- and disinformation is creating a perfect storm which is threatening democracies in Europe, in the United States and elsewhere. One thing that is occurring to me these days, and I would like to make a particular point about this, is that for a democracy to be healthy we need to be able to have a constructive difference of opinion. That's when democracy is most healthy. That is when we have a "marketplace of ideas" where we can disagree and still appreciate each other's perspectives, debate it out and combine the best ideas for our society. But what we are seeing now is a divergence of realities. This is happening for a variety of reasons - the splinternet[7] where different people and different opinions have been living in completely different ecosystems because they are fed very different information by the platform algorithms (Bakshy *et al.*, 2015, Claussen *et al.*, 2019). They watch completely different cable news networks and they read different journals and listen to different journalists. So, we have completely divergent views of reality that are created by the sociotechnical environment that we live in now. In addition, affective political polarisation and mis- and disinformation makes it extremely hard to find common ground, which means that the real benefit of democracy, i.e. the "marketplace of ideas", gets very difficult to accomplish. This is when conflict between two or more sides becomes more and more likely. We see this happening all over the world and I think that is a perfect storm that does threaten democracy.

Domalewska: Is social media a good place for a constructive difference of opinion? Social networking sites started as a forum for sharing information and exchanging opinions. In one of your studies, you showed that users expand their exposure to novel, most unique information through their weak bridging ties (distant acquaintances outside the core network that serve as bridges to other networks), which form low bandwidth and structurally diverse channels (Aral and Dhillon, 2022). But what we have now is filter bubbles and the splinternet that decrease the diversity of opinions. How can we promote this constructive dialogue in social media?

Aral: Yes, I do believe that there are ways to design social media to support a healthy and constructive dialogue. I don't believe that we are successful at that today and I think there are a number of reasons why that's the case. But I think that it certainly is possible. I think there is a need to really think about how to design and implement constructive discourse and dialogue and there are examples of that around the world. So, for instance, Twitter has focused recently on this notion of conversation health. They have implemented very small changes that are moving towards the idea of creating healthy conversations online. A couple of examples - these are not big enough, they don't go far enough but they are an example of how Twitter is thinking about how to create healthy conversations. So, for instance, they're trying to introduce critical thinking and friction into the sharing of information. For example, what you'll notice on Twitter today is that when you try to retweet an article without having clicked on the link and read the article, Twitter will ask you: "Are you sure you want to retweet that? Don't you want to read the article first?" So, this slows information down. This forces you to introduce a little more critical-thinking about what you're sharing. And it tries to short-circuit the sharing of what is sometimes mis- and disinformation, sometimes just inflammatory opinion without stopping and thinking. Another really good example is what's happening in Taiwan. The digital minister in Taiwan, Audrey Tang, has instituted a bunch of information systems that try to crowd-source ideas in ways that find consensus rather than in ways that highlight and prioritise disagreement. They highlight and prioritise areas of rough consensus. And using it as a

---

[7]*Splinternet refers to "the breaking apart of social media and digital communication networks into unconnected groups, by dimensions like political ideology or geography" (Aral, 2021).*

way to create governance, mechanisms of regulation and to pass laws in Taiwan.[8] That is a fine example of good information technology that is designed to find common ground rather than to find areas of divergence.

Domalewska: These are examples of initiatives undertaken to fight disinformation and promote constructive debates. The example of Taiwan shows that a social networking site can be used by policy-makers to actually listen to citizens. However, a growing body of research shows that polarisation in Europe is a top-down process because disagreement and division are promoted by politicians.

Aral: Yes, we obviously have this in the United States as well. Donald Trump was and is an excruciatingly polarising figure. I don't think he even tries to hide the fact that he supports polarisation. He wants to inflame conflict between the extreme right and the centre left and the left. That is part of his very productive political strategy. It can win a presidency. You see the same thing happening in France and Italy where the conservative parties are gaining lots of traction with similar strategies.

Domalewska: You have mentioned several tactics to fight disinformation. You gave the example of Twitter when discussing initiatives undertaken by social networking sites to promote a positive communication environment. How can legislation solve the problem of falsehood and manipulation but, at the same time, protect the freedom of expression and opinion?

Aral: I'll give a couple of examples. This is on my mind as we speak for several reasons. First, political speech is regulated differently than non-political speech. And the reason for that is because there is money being funnelled into political speech and there are democratic outcomes at stake. That is why laws regulate and guide what can be said in the political ad - not only how you can say it, but also you have to label it as political speech and so on and so forth. These regulations are much stricter in terms of constraining speech in the political realm than in the non-political realm. However, they have not been enforced at the same level online as they have been offline. This constraint and regulation of political speech has to be implemented with full force online, just as it is enforced offline and on television. That's number one. Number two, in the United States, the Supreme Court just last week (October 3, 2022) agreed to take up two cases and are going to look at Section 230 of the Communication Decency Act in the United States.[9] Section 230 provides a shield from civil liability for platforms being sued for what content they allow or don't allow online. The reason for that law is because if those companies are sued and are held liable for everything that is on their platforms, then they could get out of that legal obligation by just not moderating it at all. As soon as they moderate and make editorial decisions about what should be shown and what shouldn't be shown, they may be legally liable for what they allow and don't allow. According to Section 230 of the Communication Decency Act, social networking platforms are not going to be liable for editorial decisions that they make allowing some content but not allowing other content. It is user-generated

---

[8]*Taiwan platform promotes an open consultation process in which citizens are encouraged to debate and express their agreement or dissent in order to achieve consensus. Unlike sites such as Facebook or Twitter that live off the attention economy of amplifying division and emotional content, vTaiwan promotes constructive disagreement. For instance, in a heated debate about Uber regulation, the underlying concerns for safety helped to find common ground (Miller, 2020).*

[9]*Section 230 says that "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider" (United States Code, 2011). In October 3, 2022, the Supreme Court in the U.S. agreed to review two cases that concern the liability of Internet services for content that is posted: Gonzalez v. Google, about promoting terrorist content through a recommendation system and sharing ad revenue with terrorist organisations, Twitter v. Taamneh, about liability for posting terrorism-related content.*

content and platforms are not liable for that. That allows them to actually moderate the content and make decisions about what should not be shown and what should be shown. If they suddenly become legally liable for those decisions, they may just say "Well, we can't afford to moderate this content because now we can be sued". And that would make a completely free for all cesspool of information online. Now, by agreeing to review these two cases, one against Google and the other against Twitter, the Supreme Court contend that the companies in question should be legally liable for not stopping mis- and disinformation around terrorist recruitment and not stopping enough ISIS content that radicalised people, that caused terrible terrorist attacks that killed people. The Supreme Court is now going to decide how section 230 should be handled in the highest court in the United States. That could dramatically change the notion of how content is moderated online and whether platforms are liable for their moderation decisions. That opens up the question of what should we do as a society, and what should the Supreme Court and after the Supreme Court - the Congress – do, in terms of how liable the company should be for their moderation decisions and so on. I think it is reasonable that section 230 be reinterpreted to ensure the platforms are more responsible for some of the content that is being put online. But it has to be balanced with this idea of free speech online. So, for example, there have been investigations and cases into the infamous Silk Road, a dark web's website where you can buy drugs, order assassinations, and another website that was dealing with child trafficking. I think we can all agree that child trafficking should be moderated and eliminated from all platforms online. It's not always clear where the line should be drawn but there are clear things on one side of the line that we can all agree should not be allowed, which means that there should be a line. And so the debate is how we should think of drawing those lines in a reasoned way. In a way that is fair and reasonable and really does pay attention to free speech. In my book, I argue that this line has to be drawn in a very deliberate fashion.[10] It can't be a blunt instrument. You have to use scalpels not axes to draw this line. It has to be very precisely drawn and say that these are things we clearly want to forbid and the rest is free speech. Because it can be drawn in a way that is extremely hurtful to the marketplace of ideas, to free speech. We certainly don't want to have censorship, but there are things that certainly should be moderated out. And I think that platforms need to moderate their content but that having some legal responsibility for moderation is OK as long as those lines are very clearly delineated as to what exactly they should be legally liable for and everything else they should be shielded from being liable for. So this is a very difficult conversation but one that is becoming really important and a really hot topic in the United States now.

Domalewska: Moderating content is like walking on thin ice because some information can be true according to one person's political views but false or manipulation when judged by another person. An infamous example of content moderation is the Russian Federal Service for Supervision of Communications, Information Technology and Mass Media (Roskomnadzor), a federal executive body responsible for overseeing, controlling and censoring the media, including the new media. It imposed heavy fines, slowed down the activity, blocked some content and even blocked access to some social networking sites for allowing content related to the Russian invasion in Ukraine, which is deemed 'fake posts'. So, Russia is using content moderation to impose censorship and stifle free speech.

---

[10] *"When we think about how to balance free speech and harmful speech on the Hype Machine, we should consider the process before the outcome. Limiting free speech raises important ethical and constitutional questions. The case-by-case nature of speech decisions and the fact that society's norms and circumstances will likely change as technology changes suggest the importance of a deliberative process that defaults the broad protection of free speech and leaves to the courts the task of defining free speech and to the legislatures the task of circumscribing protections from civil litigation over harmful speech. Where limits to speech make sense due to the egregious nature of its potential harm in specific circumstances, the process for deciding those limits should be deliberative and representative, rather than swiftly enacted by a select few. Deliberative processes make change deliberate, as it should be for limitations imposed on free speech" (Aral, 2020, p. 319).*

On the other hand, some political parties or political figures in democratic countries are banned from some social media platforms. Should platforms make such decisions as to whether to ban a political party because of its extreme opinions?

Aral: No, I don't agree with that at all. I don't think Donald Trump should be banned from Twitter. I would air on the side of non-censorship. But I think that we have to understand that the things that we all agree on should not be part of our information ecosystem. That does not mean that moderating that content and holding platforms liable for that content opens the door or is a slippery slope to censorship. I don't believe that. I believe that we have to be able to draw those lines precisely and brightly enough so we moderate content that we know is harmful but that we allow everything else in the name of free speech and holding a bulwark against censorship. But I think that is a very difficult task. For example, there's a common phrase in the U.S. - one person's terrorist is another person's freedom fighter - which essentially means that although we can all agree in the abstract that terrorism is bad, but defining who is a terrorist and who is not a terrorist is sometimes extremely political. That's where things get very difficult.

Domalewska: Could you give some recommendation on how to protect the truth and how to improve the social media ecosystem that is rife with fake news and polarisation?

Aral: In my book, I describe the four levers that we have at our disposal that we really can turn things around. To turn around the extremely negative direction that our global information ecosystem is going in, which is creating harmful spillovers into our public health, into our democracies and into our economies. And those levers are money, code, norms and laws. Money is about the business model of social media platforms that make this information ecosystem. Right now, the platform runs on the attention economy, which means that getting people hyped up and excited and engaged is what makes profit. That makes extreme content possible because it is engaging and it spreads further, faster and deeper. We know that disinformation spreads further, faster and deeper and more broadly than the truth. Things that are reshared and clicked on and again reshared and clicked on – this is what makes profit. It's a recipe for misinformation being profitable. And there are no disincentives for that today. So, the economics of social media is one aspect. The other part of that money aspect is the idea of network effects,[11] which lead to near monopolies because there is no interoperability of access to a level playing field on which different social media services can fairly compete for customers. I sincerely hope that Web 3, which is built on decentralised applications, blockchain-based applications etc., will bring a lot more interoperability than Web 2 offered. Preventing market concentration will enable competition and create costs rather than just profitability for creating bad outcomes and content. So, the first is money, the attention economy. The second is code, i.e. the design of the platforms and their algorithms. We talked about what Twitter is doing and what Taiwan is doing. This is coding for healthy communication, coming together, reaching a consensus and finding common ground. Norms involve how we as users and citizens use these platforms in our daily lives. We have to change the way we engage with digital media. That starts with educating our children properly in digital media literacy, having courses in high school and even before high school on how to spot misinformation and disinformation and fake things online, deep fakes, and also just how to promote critical thinking, critical reasoning and digital media literacy. We as a

---

[11]*"For products or markets that exhibit network effects, the value of the product or market is a function of the number of people connecting to it. As more people use the products, its value to everyone increases (...) Network effects are like gravity. The greater the number of people on a network, the greater its 'mass,' so to speak. The greater its mass, the greater its gravitational pull. The greater its gravity, the more attraction it has to pull in new customers, and the greater the economic grip it has on preventing current customers from leaving its orbit" (Aral, 2020, p. 113).*

society have to change the norms of how to use these technologies. And the final one is regulation, which includes section 230 in the Communication Decency Act, the Digital Market Act and Digital Services Act in Europe. These are two prime examples of legislation in Europe that are moving in the direction of creating a regulatory environment that is pro-communication health and pro-democracy and pro-economic competition, which is important. So it's money, norms, codes and laws. And I really do think that we have to take all of these together. That there is no such thing as a silver bullet but together we can make meaningful changes.

Domalewska: Thank you very much for the interview.

Sinan Aral is the *David Austin Professor of Management, IT, Marketing and Data Science* at MIT, Director of the *MIT Initiative on the Digital Economy* (IDE) and a founding partner at *Manifest Capital*. He was the chief scientist at *SocialAmp*, one of the first social commerce analytics companies (until its sale to Merkle in 2012) and at *Humin*, a social platform that the *Wall Street Journal* called the first "Social Operating System" (until its sale to Tinder in 2016). He is currently on the advisory boards of the Alan Turing Institute, the British National Institute for Data Science in London, the Centre for Responsible Media Technology and Innovation in Bergen, Norway and C6 Bank, one of the first all-digital banks of Brazil.

His research has won numerous awards including the Microsoft Faculty Fellowship, the PopTech Science Fellowship, an NSF CAREER Award, a Fulbright Scholarship, and the Jamieson Award for Teaching Excellence (MIT Sloan's highest teaching honour). In 2014, he was named one of the "World's Top 40 Business School Professors Under 40." In 2018, he became the youngest ever recipient of the Herbert Simon Award of Rajk László College in Budapest, Hungary. In the same year, his article on the spread of false news online was published on the cover of *Science* and became the second most influential scientific publication of the year in any discipline, and his TED talk on "Protecting Truth in the Age of Misinformation," which received over two million views in nine months, set the stage for today's modern solutions to the misinformation crisis. Sinan's first book, *The Hype Machine*, which was named a 2020 Best Book on Artificial Intelligence by *WIRED*, a 2020 Porchlight Best "Big Ideas and New Perspectives" Book Award Winner and among the Best New Technology Books and Best New Economy Books to Read in 2021 by Book Authority, became an instant classic.

References

Aral, S. (2020) *The hype machine: How social media disrupts our elections, our economy, and our health—and how we must adapt.* New York, NY: Currency.

Aral, S. (2021) [Twitter] 22 January. Available at: https://twitter.com/sinanaral/status/1352657222279557125 (Accessed: 29 October 2022).

Aral, S. (2022) 'Ukraine is winning the information war', *The Washington Post,* 1 March. Available at: https://www.washingtonpost.com/outlook/2022/03/01/information-war-zelensky-ukraine-putin-russia/ (Accessed: 29 October 2022).

Aral, S. and Dhillon, P.S. (2022) 'What (exactly) is novelty in networks? Unpacking the vision advantages of brokers, bridges, and weak ties', *Management Science*. doi: 10.1287/MNSC.2022.4377.

Bakshy, E., Messing, S. and Adamic, L.A. (2015) 'Exposure to ideologically diverse news and opinion on Facebook', *Science*, 348(6239), pp. 1130–1132. doi: 10.1126/SCIENCE.AAA1160.

Berger, J. and Milkman, K.L. (2012) 'What makes online content viral?', *Journal of Marketing Research*, 49(2), pp. 192–205. doi: 10.1509/JMR.10.0353.

Boxell, L., Gentzkow, M. and Shapiro, J.M. (2017) 'Greater internet use is not associated with faster growth in political polarization among US demographic groups', *Proceedings of the National Academy of Sciences of the United States of America*, 114(40), pp. 10612–10617. doi: 10.1073/PNAS.1706588114.

Claussen, J., Peukert, C. and Sen, A. (2019) 'The editor vs. the algorithm: Targeting, data and externalities in online news', *SSRN Electronic Journal*, 5 June. doi: 10.2139/SSRN.3399947.

Collis, A., Garimella, K., Moehring, A., Rahimian, M.A., Babalola, S., Gobat, N.H., Shattuck, D., Stolow, J., Aral, S. and Eckles, D. (2022) 'Global survey on COVID-19 beliefs, behaviours and norms', *Nature Human Behaviour*, 6(9), pp. 1310–1317. doi: 10.1038/s41562-022-01347-1.

Golding, P., Sousa, H. and van Zoonen, L. (2012) 'Trust and the media', *European Journal of Communication*, 27(1), 3–6. doi: 10.1177/0267323112439366.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N. and Westwood, S.J. (2019) 'The origins and consequences of affective polarization in the United States', *Annual Review of Political Science*, 22, pp. 129–146. doi: 10.1146/ANNUREV-POLISCI-051117-073034.

Kim, J.N. and Gil de Zúñiga, H. (2020) 'Pseudo-information, media, publics, and the failing marketplace of ideas: Theory', *American Behavioral Scientist*, 65(2), pp. 163–179. doi: 10.1177/0002764220950606.

Miller, C. (2020) 'Taiwan's crowdsourced democracy shows us how to fix social media', *We are not divided*. Available at: https://wearenotdivided.reasonstobecheerful.world/taiwan-g0v-hackers-technology-digital-democracy/ (Accessed: 29 October 2022).

Norris, P. (2017) 'Is Western democracy backsliding? Diagnosing the risks', *The Journal of Democracy*. HKS Working Paper No. RWP17-012. doi: 10.2139/ssrn.2933655.

United States Code (2011) *Telegraphs, telephones, and radiotelegraphs*. 2006 Edition, Supplement 5. Available at: https://www.govinfo.gov/app/details/USCODE-2011-title47/USCODE-2011-title47-chap5-subchapII-partI-sec230/context (Accessed: 21 October 2022).

Vosoughi, S., Roy, D. and Aral, S. (2018) 'The spread of true and false news online', *Science*, 359(6380), pp. 1146–1151. doi: 10.1126/SCIENCE.AAP9559.