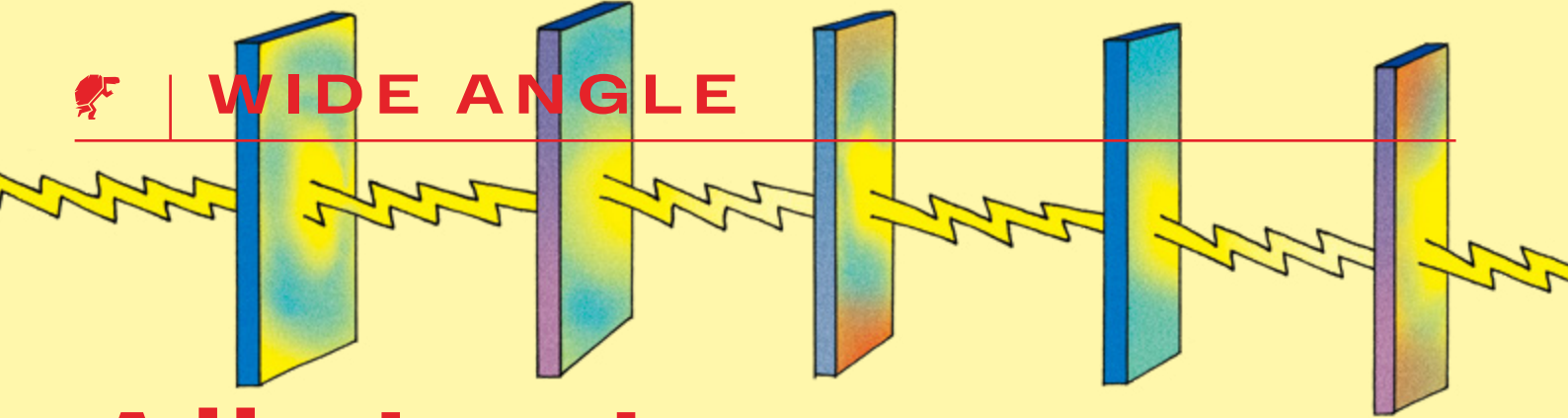




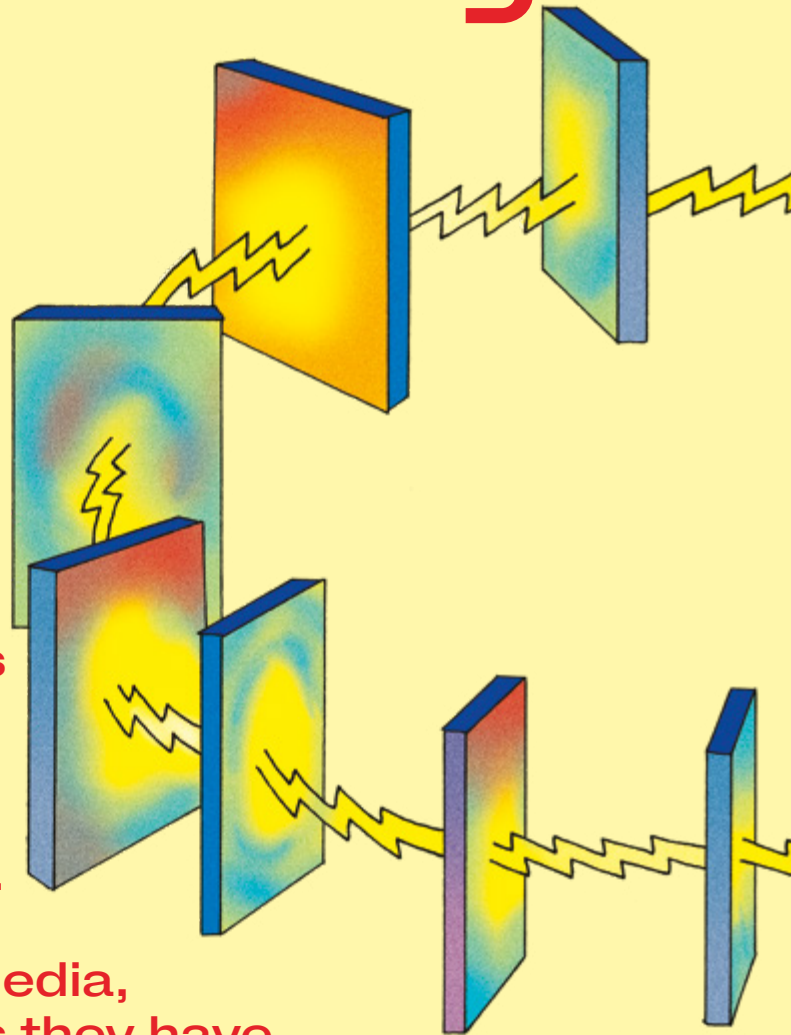
WIDE ANGLE

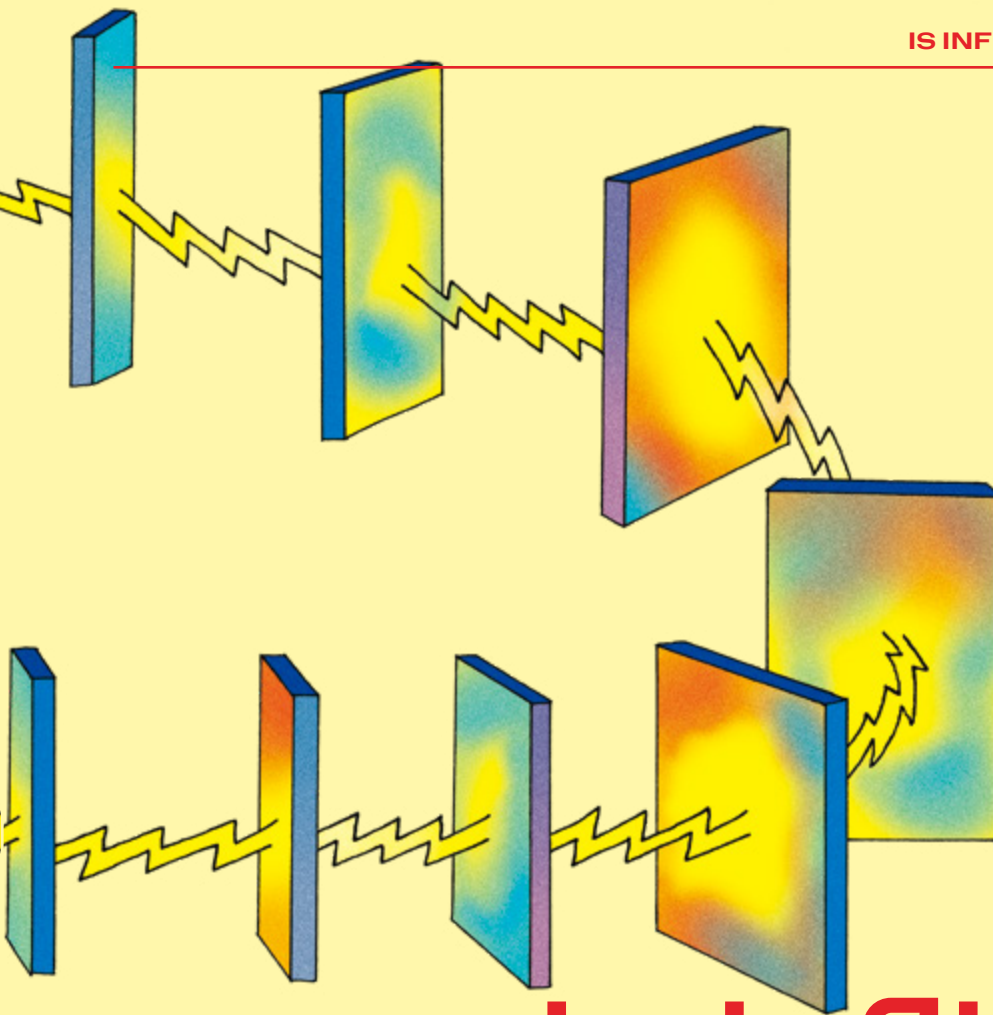


All the harm virality

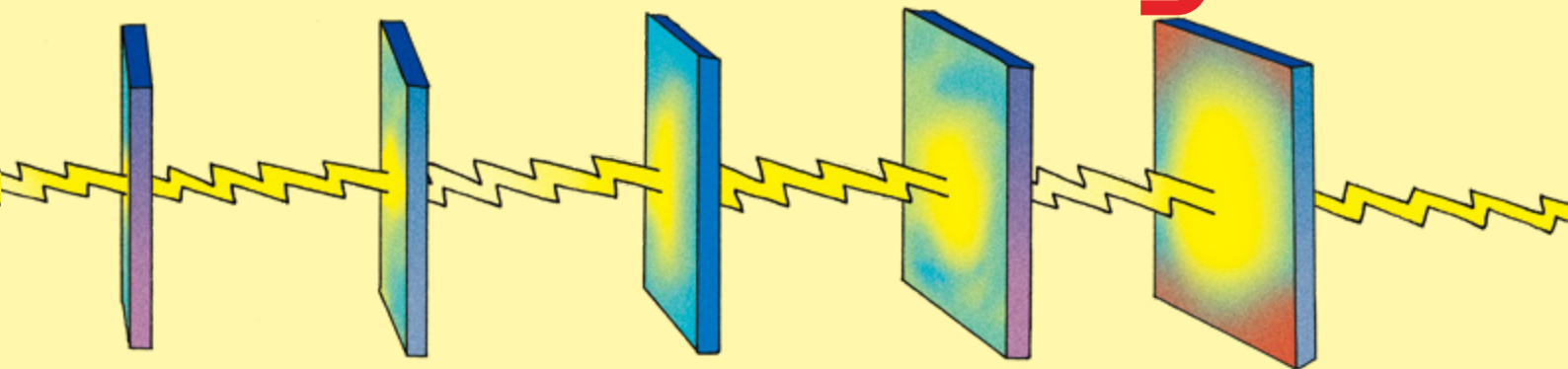
By Dominique Boullier

Regularly accused of being responsible for the deterioration of public discourse and even the destabilisation of democratic regimes, social networking platforms act as media in their own right, even though they continue to operate largely outside any legal framework. They are also seen to undermine the traditional media, whose advertising revenues they have siphoned off, and above all they boost a viral effect that weakens content through propagation chains. Dominique Boullier examines the mechanisms involved.





is inflicting on democracy



Dominique Boullier is university professor emeritus and researcher at the Centre d'études européennes et de politique comparée (CEE) (Centre for European Studies and Comparative Politics). A sociologist and linguist, he explores digitally-shaped behaviour. His current research focuses on the technical and institutional conditions for survival in digital environments.



No democratic regime appears to be safe from destabilisation. This observation might suggest looking for culprits among well-known common enemies, including terrorist movements and dictatorial regimes of all stripes. It is more difficult to identify internal vulnerabilities shared by all democracies, given their different political and economic conditions. However, since the mid-2010s, a single phenomenon has affected all political regimes: the profound change in the media landscape. The proliferation of media, their increased dependence on large financial groups, their internationalisation, and the massive domination of the moving image over the written word, have changed the status of the previous guarantors of democratic order, namely the press and journalists.

The detrimental effects of advertisement monetisation

The most radical challenge to this status has undoubtedly come from the pressure exerted by social media. Although they publish content, social networks are not subject to any of the responsibilities borne by traditional media, which are strictly regulated. This advantage stems from various pieces of legislation. Since the United States' Clinton-Gore administration put in place the Communications Decency Act (CDA) of 1996 when the public web was transforming into a commercial web, legislation has granted host status to what were then known as internet service providers. Thanks to this status, further entrenched by the European directive of 2000 on electronic commerce and by the French law of 2004 on trust in the digital economy (LCEN),

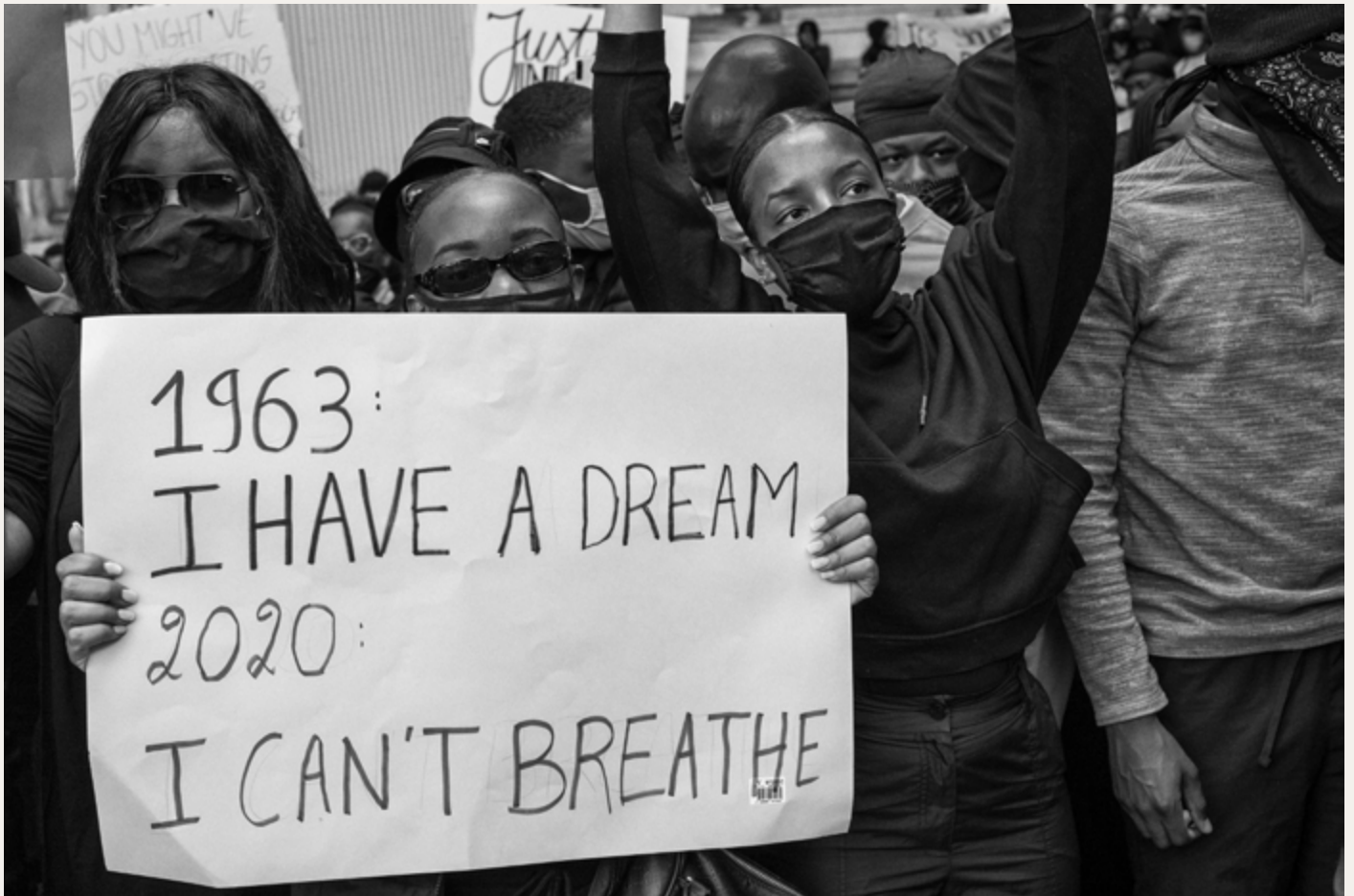
social networks have limited responsibility for the content published by their users. While such protection made sense during the internet's expansion phase, it has wrought the unsustainable development of social media, which now follow an editorial policy of deleting content and accounts. Most of today's problems are attributable to this formal flaw inherited from the CDA, as well as the challenge of fighting an all-powerful, profitable and uncontrollable system.

Another change occurred in 2008–2009, when social media, which at the time were essentially convivial, began to monetise their users' activities by selling advertisement placements to advertisers in the name of participative Web 2.0 culture. In 2016, Google (mainly via YouTube) and Facebook captured 75 per cent of the online advertising market, threatening its economic equilibrium. As they are not subject to laws that apply to other media, social media have become major advertising platforms and sources of unprecedented revenue. They have acquired a position that makes them financially omnipotent in innovation choices and culturally omnipotent in public life. This digital Wild West results from the liberal ideology of business at all costs and the systemic benefits it brings to brands and investors. But it also draws on support from audiences, including those who act as vectors of political communication.

Initial disillusionment came in 2018 with the misuse and abuse of personal data during the Facebook-Cambridge Analytica affair, which revealed a lack of protection for personal data. Governments have been very slow to respond to these risks. Although the European Union's

THE FACEBOOK CAMBRIDGE ANALYTICA AFFAIR

In 2018, whistle-blowers revealed that between 2014 and 2016 the personal data of 87 million Facebook users had been exploited without their consent by a British political marketing company, Cambridge Analytica. This personal information was aggregated with other data and used to send targeted messages on social media according to psychological profiles based on a model originally developed by an academic using tests taken on Facebook. In 2016, this data exploitation may have tipped the scales towards the Brexit vote in the United Kingdom and the endorsement of Donald Trump in the United States Republican primaries. The scandal led to several parliamentary enquiries in the United States and Europe. As a result Facebook was fined USD 5 billion for failing to protect its users' privacy, and Cambridge Analytica was forced to close operations. The affair also led governments and Europe to strengthen data protection regulations.



General Data Protection Regulation (GDPR) was passed in 2016 and implemented in 2018, it was not until 2022 that the European Digital Services Act (DSA) and Digital Markets Act (DMA) were passed. Various regulatory models continue to ignore the advertisement issue and related abuses in capturing the attention of audiences. In reality, neither Europe, the United States, nor any country in the rest of the world, has taken regulatory measures to address the virality phenomena encapsulated in the economic model of the platforms and the algorithms that run them.

The spreading culture of virality

The disease gnawing at democracies is viral. The quest for visibility has become a permanent reputational issue that appears everywhere. In particular, virality has taken hold in the world of finance, where it is pervasive. Some players now engage in a deleterious speculative game consisting of manipulating the perceptions of other investors with the goal of provoking a market reaction and capitalising on available liquidity.

They sometimes go so far as to infest the market with false promises to buy, which they withdraw before the end of the day on high-frequency trading markets. Financial market players are no longer on the lookout for the fundamentals of a stock or a company, but rather for signals propagating a belief, expectations or anticipations about a given stock that will enable them to speculate upwards or downwards. Thus, brands no longer depend on their results and sales figures so much as on investor reactions. Statistics on content views and shares (analytics) on platforms feed the fiction of the brand's relationship with its customers and its potential value (goodwill).

All this requires mobilising traces of user activity and engagement. To this end, algorithms favour content that elicits ever greater responsiveness. The 'novelty score' of content – an indicator developed by Soroush Vosoughi, Deb Roy and Sinan Aral – is essential to capturing attention. Most fake news does so by hewing to the unbelievable, divisive, unprecedented, funny, or sinister. Reactivity breeds virality, stimulated by

Demonstration in support of the Black Lives Matter movement, Brussels, 7 June 2020.



the algorithms and properties of the platforms, such as the trending topics that X/Twitter provides to all its users. These serve as indicators not of the overall popularity of a hashtag, but of the acceleration of activity around it (that might drop back down within an hour). All other media follow these virality indicators to avoid missing out on a topic that is garnering attention. Journalists suffer from a syndrome akin to FOMO (fear of missing out). As a result, any other topic that is more complex and less shocking, and that requires more time to explain or is likely to fuel a more documented controversy, is relegated to the bottom of the list. Its interest and visibility decrease as the relative virality of other topics and posts increases.

Internet users themselves, be they robots or not, quickly learn to format their own publications in a way that is conducive to going viral. The collective learning of these attention-grabbing publication formats is now occurring at a very early age among younger generations on platforms they are attracted to, such as Snapchat and TikTok. All the participants in public debate feel forced to adapt. They think they are reaching different audiences, when in reality they are condemned to distorting their own message and losing their uniqueness to become a drop of water in the viral flows propagated by the platforms according to their own criteria. The general dumbing down and polarisation of political content in public debate result from a phenomenon that affects all media, including infotainment programmes broadcast by the mass media, such as ‘Quotidien’ (Today) on TMC, ‘Touche pas à mon poste’ (Don’t Touch My Post) on C8, and repetitive news channels, also known as 24-hour news channels.

A self-replicating media ecosystem

If we are to fully understand the continued deterioration of the public arena, it is essential to consider the historical and functional propagation chain in its entirety. It is worth breaking down to highlight all the guardrails that could have been implemented, and still could be, to stem the deterioration. Each element constitutes a condition (and not a cause) enabling the next. Their succession forms a very robust whole that would be much more difficult to challenge than a chain where only the last links were addressed.

This chain unfolds from end to end in the following way:

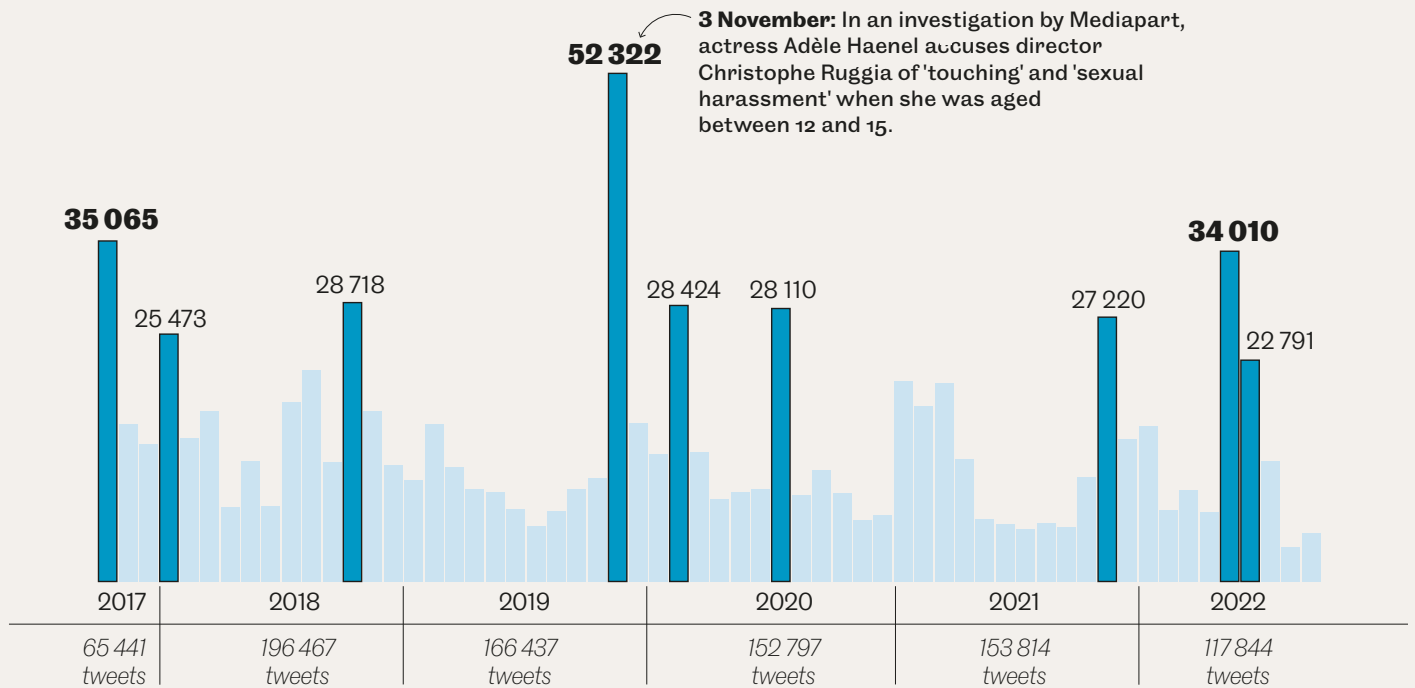
Platforms, hosts and non-media – Advertising and monetisation of the content itself – Algorithm design maximises engagement, captures attention and maximises virality – Adoption of the same rhythm and formats by the mass media – Uniformity of content designed for virality (short format, video, shock effect, catch phrases, culture of confrontation) – Platform AI systems gather personal data and learn profiles to better capture attention – Refusal to truly moderate, plus the weakness of structural regulation – ‘Self-replicating environment’ amplified by generative AI – Triumph of the fake at every level – Society of widespread distrust.

It is possible to act on each of these elements. To do so requires consideration of the self-replicating media environment that has developed and swept away all the filtering power of mass media and journalists. Equally important to bear in mind are the producers of propagandist content, who are rightly under increased scrutiny. These content producers cannot be countered without understanding that they have themselves adapted to the self-replicating environment. A case in point is the Internet Research Agency (IRA), a Russian agency created in 2013 in St Petersburg by the oligarch Yevgeny Prigozhin. From the outset, rather than disseminate sharp and targeted propaganda messages, it proceeded as an agency of ‘propagationists’, that is, geeks able to mobilise armies of trolls and robots to destabilize democracies by simultaneously launching variants of messages on the web that were very different from one another and shared only one feature: to provoke virality through the shock effect. Any of these variants might gain the upper hand depending on different properties and a particular situation. Meanwhile, the cost would remain minimal since, by definition, internet users themselves generate virality simply by sharing, posting, liking or commenting, without any barrier.

Consequently, any strategy to counter disinformation must be combined with a strategy to break the chains of propagation. Intentional propagation can only succeed because some messages have intrinsic power – often stemming from semiotic variations (see example

The #metoo hashtag on X/Twitter in France, 2017-2022

Number of tweets with the #metoo hashtag



Infographic: *Le Monde*

Source : Visibrain

opposite) linked to chance, as in all evolution – and because the algorithmic architectures of the platforms are designed to encourage this dissemination and virality. The world of social media, which have become merchants by accepting advertisements and monetising data, is economically and technically (and sometimes morally and politically, in the case of Elon Musk) designed for high-frequency propagation. Of course, this is not the case for all contemporary digital devices, and platforms such as Mastodon, Twitch and especially Wikipedia offer examples of a different approach that curbs virality with different, collectively controlled technical solutions. In any case, claims to be building European champions – such as those made by the founders of the French search engine Qwant and those now related to AI – are doomed to impotence vis-à-vis the economic clout and user bases of American and Chinese platforms. What’s more, these ambitions are counter-productive. Other technical and economic architectures should be encouraged to better align with European values of law, privacy, science, adversarial debate, verified information and informed debate in democracies.



POLYSEMIOUS FROG
 Pepe the Frog is the hero of the *Boys' Club* comic strip created in 2005. He plays the role of a teenage video game fan. Since 2008, the character has been the subject of numerous internet memes. One of them, sometimes sporting Trumpian hair, has been used by the far right in the US to advance supremacist ideas. In China, a tearful version served as a rallying point for opponents of the regime, before becoming one of the symbols of the Umbrella Movement in 2014 and 2019 in Hong Kong.

Breaking the chains of contagion and building protection from virality

Pending the implementation of fledgling initiatives and the emergence of public backing to support them without controlling them, how can these chains of propagation be broken? As long as social media are not considered to be real media and advertising on them is not banned, the virality at the heart of their model will persist. Even if they maintain their status as hosts and homes to advertising, many regulatory measures are still conceivable.

In a policy brief entitled *Social Media Reset* published in 2024 by the Digital, Governance and Sovereignty Chair at Sciences Po, I proposed 29 measures designed to restore a liveable media environment. One of them recommends that each internet user’s online activities scores be permanently displayed on their screen using a dashboard that cannot be deactivated. Analogous to a steering instrument that is required to control the vehicle (and similar to many video game displays), this dashboard would indicate the following elements: length of connection, number of posts, number of likes, number of shares, and



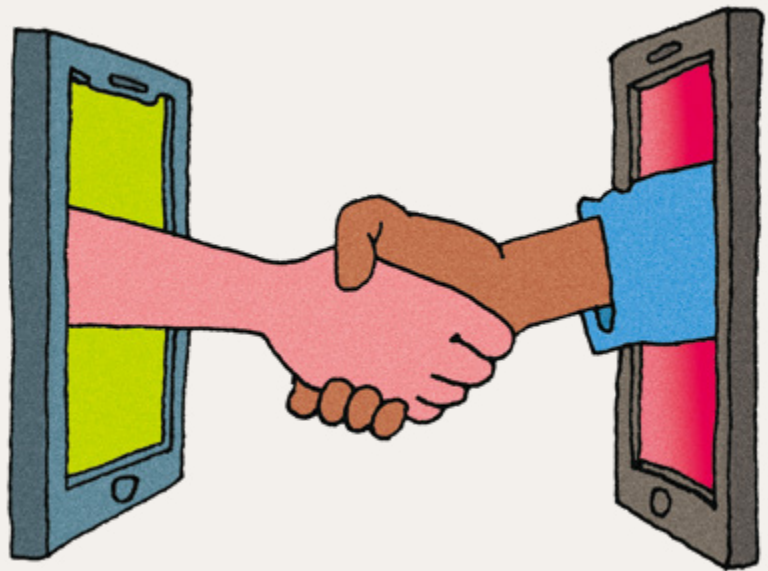
number of comments over a 24-hour period. A threshold could be set to trigger alerts in the event of excessive propagation. These alerts would help people transition from a spontaneous cognitive position to one of self-control, reflection and explicitness, thus disrupting the automatic behaviour that platforms are constantly enabling (such as one-click purchases in disregard of informed consent). The system would only address *reach* – that is, the number of people seeing your content – activated by a viral mechanism through pure reflex; it would not affect content or challenge *free speech*, but would only defer the sharing of content. And accounts that are very large replicators could be blocked or even deleted, while illegal content could be tracked down. It would be more a matter of setting in motion a system of responsible self-monitoring among internet users, before introducing collective mental speed regulators when certain high-risk conditions are met.

The digital space involves the same type of ‘multiplayer action’ as car traffic, as described by Pierre Livet and Laurent Thévenot. It is neither ‘collective action’ nor ‘joint action’, but a situation where people act separately and simultaneously in the same environment. Inherent in this type of situation is the fact that none of the players is aware of the scale of the interactions taking place, and each tends to think that his or her own actions are of no consequence, even though, in the case of digital networks, each internet user contributes to virality and to alerting the entire system of collective attention. Digital information flows have much to learn from road regulation if they are to become liveable again. Just as on the road, it will no doubt be necessary, at certain times or for certain accounts, to take restrictive measures, setting a maximum threshold for reactivity. This will provide both a means of alerting the public to their own behaviour and its aggregate effects, and a slowdown mechanism that can only benefit public life.

An independent body to collect and measure platform traces

Implementing this type of policy hinges on access to data, which is generally limited to behavioural traces on each platform. This requirement should be included in the specifications of all platform operators in a given territory. A trusted

The digital space involves the same type of ‘multiplayer action’ as car traffic, where people act separately and simultaneously in the same environment.



THE EUROPEAN REGULATORY SYSTEM

- **General Data Protection Regulation (GDPR)**
Adopted in 2016, applicable from 2018

Strengthens the protection of European citizens’ personal data and harmonises the rules governing the management of this data by public and private organisations.

- **Digital Markets Act (DMA)**
Adopted in 2022, applicable from 2023

Imposes rules on digital platforms to combat anti-competition practices.

- **Digital Services Act (DSA)**
Adopted in 2022, applicable from 2024

Imposes rules on digital platforms to hold them accountable and to combat the dissemination of illegal or harmful content and the sale of illegal products.

- **European Artificial Intelligence Act (AI Act)**
Adopted in 2024

Defines three categories of AI applications: unacceptable risk applications (e.g. government-run social rating systems, such as those used in China); high-risk applications (e.g. a curriculum vitae classification tool); and risky applications.



'MeToo' mural on the wall of a girls' school, Parmer, Rajasthan, India, August 2024.

third party such as the ones that exist for mass media (for example, Médiamétrie in France) would be needed – that is, a body that is neither judge nor party. It would be responsible for gathering traces at the source, analysing them and sharing them with researchers (given that the information is of public interest) and with the parties involved in all the advertising revenues generated by the platforms, namely the advertisers. The latter are currently billed in a very opaque way (auction system) for placements for which the real impact on buying behaviour is never measured (since it is the brands' reputations that count for investors), hindering them from more fully understanding the engagement of their own customer community. Tim Hwang, former director of the Harvard-MIT Ethics and Governance of AI Initiative, a philanthropic research fund, called this fool's deal 'the online advertising bubble'. The market needs to be regulated. Only a deep understanding of the mechanisms of virality, as formalised in my book *Propagations, a New Paradigm for the Social Sciences*, combined with strong political will, will enable the re-establishment of a liveable public space.

REFERENCES

- Boullier, D., *Comment sortir de l'emprise des réseaux sociaux. Le réchauffement médiatique*, Paris: Le Passeur éditeur, 2020.
- Boullier, D., 'Puissance des plateformes numériques, territoires et souverainetés', *Research Paper*, Digital, Governance and Sovereignty Chair, Sciences Po, April 2022.
- Boullier, D., *Propagations. Un nouveau paradigme pour les sciences sociales*, Paris: Armand Colin, 2023.
- Boullier, D., 'Social Media Reset. Redesigning the Infrastructure of Digital Propagation to Cut the Chains of Contagion', *Policy Brief*, Digital, Governance and Sovereignty Chair, Sciences Po, June 2024.
- Livet, P. and Thevenot, L., 'Les catégories de l'action collective', in André Orléan (ed.), *Analyse économique des conventions*, Paris: PUF, 1994.
- Vosoughi, S., Deb, R. and Sinan, A., 'The Spread of True and False News Online', *Science*, 359 (6380), 2018, pp. 1146–1151.