11. Indentification of potential risks related to the Future of digital media

Editor and corresponding author: Víctor Rodríguez Doncel
Authors (alphabetical order): David Chavalarias, Caroline Datchary, Dana Diminescu, Jean Lassègue, Quentin Lobbé, Víctor Rodríguez-Doncel, Fabrizio Silvestri

Keywords: trends in digital media; risk management; risk assessment; risk mitigation; long-term threats

Rationale

Previous chapters have considered the impact of current digital media on human well-being and outlined an associated research agenda. This chapter completes the picture with an attempt to identify effects of media that, as of today, have not yet been fully materialized, but are likely to appear. Trying to foresee these future effects using risk-analysis tools is of the utmost importance, for they will broadly impact media users and the society at large.

The methodologies to carry out this analysis should be object of research and they will be explored in **Topic 1** in this chapter. Although transformations are happening at an unprecedented pace, anticipation of likely future trajectories is possible at least for the short term future. The most notable of these risks relate to how groups form online and their associated social dynamics (**Topic 2**), or how social processes affect the public perception of risks (**Topic 3**). Finally, the most difficult risks to be evaluated and anticipated are related to global, long-term and systemic problems, overviewed in **Topic 4**.

We can only speculate about the future development of new media technologies. However, extrapolation of current trends suggests the near-term future will see more active social media users (reported to be more than 4.5 billion as of 2022), spending more time online (currently more than six hours per day) and pervading more professional activities beyond leisure [97]. Image and video based social neworks such as TikTok and Instagram, currently the fastest growing networks [93], may continue to gain momentum. Voice search and the interaction through

smart assistants may also continue growth. However, other changes are to be expected beyond the extrapolation of current trends. Major transformations could take place due to regulatory changes (e.g. the European Digital Identity Wallet), technological developments (new disruptive technologies such as ChatGPT may emerge) or even have a black-swan nature (e.g. remote education and working brought about by COVID-19).

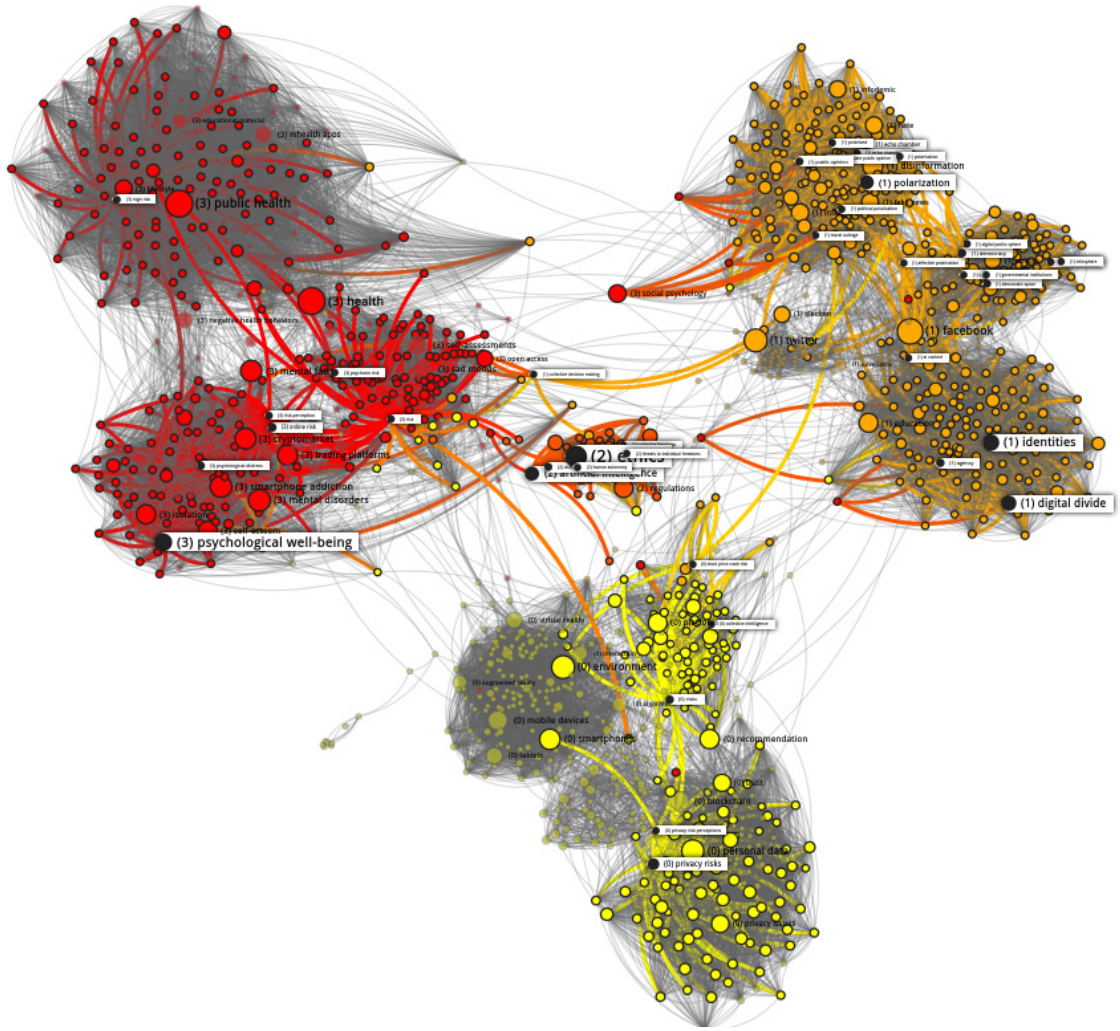

Figure 11.1: State-of-the-art domains concerned by this chapter. Map made with *GarganText*. Interactive map available online.

Topic 1. Methodologies for risk assessment in future digital media

Keywords: foresight analysis; ethical foresight analysis; risk assessment; risk management; incident databases

There is a generally acknowledged need to assess the desirability of new and emerging technologies early in their development. It is better to anticipate developments when they are still malleable, even if at this stage the future is still uncertain. Foresight Analysis is a methodology used since the 1950s for predicting the outcome of potential policy decisions, emerging technologies and artefacts, as well as economic and societal trends. Methodologies for doing Ethical

Foresight Analysis have also been proposed, with the goal of identifying and predicting the most salient ethical issues likely to arise from new technological artefacts, services, and applications [25, 34, 87]. These techniques can also be used to anticipate potential risks in future digital media, but they need a place on their own in the research agenda.

Key challenges and questions

▪ **Challenge 11.1 — Foresight analysis methodologies.** Specific methods and techniques must be used to identify individuals and groups affected by future digital media developments, to anticipate and predict the risks affecting these individuals and groups and also to plan mitigation measures. Research questions include: How should these methodologies be adapted to the specific case of future media? How can we integrate both online and offline methodologies? How can we best anticipate unlikely technological, social or political developments? Are standard risk management procedures (such as those in ISO 31000) also applicable in this domain?

▪ **Challenge 11.2 — Datafication of digital media-related incidents.** In the last few years incidents related to AI systems have been systematically collected in databases. One such collection, the Artificial Intelligence Incident Database (AIID) [94], classifies incidents collected from the news with a taxonomy of seventeen harms or near-harms categorised by five severity levels. Equivalently, databases of harms related to digital media might be collected. These databases may contain not only news but also other resources such as biographical interviews, data related to self-confrontation psychology experiments, field knowledge and other kind of ethnographical research data. The availability of this connected network of datasets may lead to the creation of knowledge graphs and social digital twins [21], allowing statistical analysis of individuals, groups, media and their relations. Research questions are: Can such deeply qualified datasets be used to reliably infer potential risks? Is simulation of different scenarios (what-if) possible? Other research questions around datasets of incidents related to future digital media include: How can we categorise risks related to digital media? Which databases can be systematically collected for the rigorous study of risks? Which features should be reflected in the media-related incident databases?

Topic 2. Risks associated with group formation and dynamics

Keywords: echo chambers; filter bubble; polarization; human autonomy; collective decision making; digital divide; digital literacy.

Digital media change the way we make social ties and influence one another, profoundly altering the processes of social groups formation and the circulation of information within and between these groups. Backstrom et al. [88] identified the structural determinants in social networks that most influenced these processes in the early 2000s, but the significance of the media and the formed groups has changed since then, and a broader analysis is now necessary.

New areas of study such as the sociology of the internet or the digital anthropology still have to explore the transition from a group-based to a network-based society that is decoupling community and geographic proximity, and thus requiring new understanding [17]. Exclusion from groups how has more severe consequences, and new digital divides require more attention. Meanwhile, being too strongly influenced by a group leads to opinion polarisation, a likely effect of group-formation dynamics that locks media users into closed communities. The global reach of the new media endangers diversity of thoughts, freedom of expression and ultimately the health of collective decision making processes.

Key challenges and questions

▪ **Challenge 11.3 — Digital divides.** New media technologies create digital divides with disparities in access, usage and influence, these divides driven by a variety of factors. These include sociodemographic and socioeconomic characteristics, personal elements, types of technology, degree of digital training, rights, infrastructure, among others [35]. Whereas existing divides have been explored, future divides may emerge from differences in algorithmic awareness or data inequalities. Identifying these new divides is a key research challenge, as is learning to identify the groups and collectives that will be most affected, and the likely consequences of group marginalisation. Other questions might be related to the mitigation measures: how can digital media literacy be improved? How can we best teach such literacy? What are the conditions for upskilling segments of the population that are currently most in need of education (e.g. seniors)?

▪ **Challenge 11.4 — Echo chambers and polarisation.** The Web 2.0 (social web) era has reinforced phenomena such as echo chambers and filter bubbles [14]. Individuals in echo chambers tend to consume and spread information aligned with their pre-existing beliefs. Yet algorithms in social media also favour equally aligned content, entering into vicious circles that lead to the polarisation of public opinion, with black-or-white opinions and beliefs. The possible perspectives and horizons of individual thought become limited [61]. Whereas this is a current issue, we need research on the long- term evolution of such dynamics. What are the effects of less randomness in encounters? How to prevent digital media bubble groups to evolve into digital ghettos?

▪ **Challenge 11.5 — Collective decision making.** New media exert a strong influence on collective decision making. The following questions remains open: Is there a way to maintain democratic control amidst ever-growing influential algorithms? Is it possible to avoid the creation of new digital feudalities in which platforms or individuals exert undue influence? Is there a way to preserve identification continuity in the process of collective action under the deployment of new technologies?

Topic 3. Social construction of risks in public opinion

Keywords: social construction; public opinion; hyper-reality.

The concepts populating the minds of the individuals of a society can be created or modified by social engineers through media technologies. Intersubjectivities can increasingly detach the collective imagination from the actual world. Something to explore is how this divergence between the real and the conceptualized becomes larger with the new media technologies, amplifying effects on human well-being. Digital media empowers agenda setting, or the ability to shape public opinion by determining which issues are given the most attention. It can be used in nefarious ways. Examples are the artificial construction of irrational fears in order to establish controversial and questionable policies, and the actions to foster moral panic or bellicose attitudes in international crisis situations.

Key challenges and questions

▪ **Challenge 11.6 — Moral panic.** The widespread feeling of irrational fear of a person, group or idea can be artificially triggered in online social networks. Online behaviour – violent attitudes, for example – can also spill over into the off-line world, and in this way, the shape of the public sphere can be manipulated. We have little fundamental understanding of the nature of the relevant dynamics, their relation to the topology of social networks; and whether moral

panic appears spontaneously or is induced by interested parties.

▪ **Challenge 11.7 — Human autonomy.** The application of artificial intelligence and machine learning algorithms to deliver personalized advertisements to individuals based on their digital footprint can become a form of manipulation. It may provoke threats to human autonomy, changing the course of elections or immersing individuals into echo chambers. The future consequences of increasingly powerful algorithms remains an open research area. How will AI-mediated communications influence human autonomy?

▪ **Challenge 11.8 — Erosion of individual identity.** For most of human history, individuals could take for granted the public validity of their individual identity. But emerging deepfake algorithms, AI-generated text and other AI techniques will democratise technologies for computers to impersonate humans or for people to impersonate others. The value of public identity could be partially lost. We must identify the immediate risks of impersonation, as well as their enabling technologies [90]. On the other hand, the implementation of measures to certify identity may lead to an excess of control threating again individual freedom. Finding a balance between these competing risks (for example in the metaverse) remains an open challenge. What are the legal and technological instruments to fight impersonation? How to empower individuals to control their own virtual identities and safeguard their biometric data?

Topic 4. Systemic, global and long-term risks

Keywords: systemic risks; global risks; long-term risks; digital preservation.

The global character of digital media, the lack of technodiversity, and the possibility of a doomsday invention [31] amplified by new digital media magnify the impact of global, systemic and long-term risks. The existential dependence of modern societies on electrical energy has been extended to telecommunications – a sustained internet failure would lead to catastrophic consequences. Whereas critical infrastructures will not depend on social media, the malfunction of the latter or their malicious operation may in the future lead to equally devastating results on society. The preservation of digital goods might be at risk, and so are democracy and public institutions.

Key challenges and questions

▪ **Challenge 11.9 — Digital preservation.** Proprietary social platforms are not open and are therefore beyond the scope of Web archive initiatives. Also, if efforts to decentralize the Web do not succeed, information archives might be in the hands of few entities, and so come to represent single points of failure. Web archives might in these cases be open to manipulation, biased or of limited diversity (socially, politically, geographically, etc.). A key challenge is to determine the means for granting web historians the digital preservation of the collective digital heritage. Should political, technological or legal instruments be deployed to grant the preservation of digital goods?

▪ **Challenge 11.10 — Democracy and public institutions.** Many challenges can be explored in relation to reinforcing democratic values and strengthening public institutions. These research avenues include methods to mitigate the risks that future digital media pose to democracy through generating unrest, instabilities, polarisation, or changing the relation between societies and their environment. Also, determining the extent of public intervention in the new digital media remains an open challenge, as well as determining the nature and extent of the control instruments (legal, technological, organisational).

▪ **Challenge 11.11 — Psychological well-being.** Psychological well-being is largely determined by self-perception. But self-perception has been dramatically reshaped in the online world, where different digital identities are created and evolve under the influence of external agents. A key challenge is to identify means to decelerate and to manage personal continuity and autonomy in ephemeral groups. How will future media technologies change our self-conception, our conception of reality and our interactions with reality? What is the precise extent of the onlife [76] in psychological well-being?

▪ **Challenge 11.12 — AI generated content.** Large Language Models and new image generation algorithms such as stable diffusion enable the generation of high-quality text and images. This AI generated content, possibly used for behavioural change purposes, may dominate the new media landscape and have incalculable consequences. This risk is particularly acute, if algorithms are iteratively refined – ChatGPT has been trained to improve by interacting with users by means of Reinforcement Learning techniques. An infosphere populated by interacting automata (inforgs of artificial nature) will irremediably imply the redefinition of what we are, repositioning the place of humans before the world [75] and creating a new sort of risks and opportunities that have to be researched.