Weak Systems
Unveiling the Vulnerabilities of Digitization

Paolo Bory
Politecnico of Milan

Philip Di Salvo
London School of Economics and Political Science
Università della Svizzera italiana

Abstract: This article discusses digitization weaknesses, biases, and malfunctions to challenge popular, almost hegemonic visions of contemporary technologies. By focusing on examples provided by recent mediated cases, controversies, and critical research about biases, we aim to propose an analysis of anything digital starting from its vulnerabilities, to look beyond polarized deterministic views, both optimistic and pessimistic. The article generates from the thematic track: "Weak Systems. Exploring bias, bugs and the vulnerability of digitization" that took place at the VIII STS Italia Conference. The panel brought together scholars from different backgrounds, including STS, history of technology, sociology of communication and critical data and media studies to discuss instances of technological weaknesses in various contexts. The article sums up some of the panel takeaways and pleas for a cooperative and interdisciplinary effort focusing on "weak systems".

Keywords: Digital vulnerabilities; infrastructures; artificial intelligence; cybersecurity; critical data studies.

Submitted: November 15, 2021 – **Accepted:** January 15, 2022

Corresponding author: Paolo Bory, Politecnico di Milano, Dipartimento di Design, Via Durando 10, Milano, Italy. Email: paolo.bory@polimi.it

1. Introduction: From Powerful Systems to Weak Systems

This article considers popular narratives of digital technologies, their origins and rationales aiming to advance a critical take on how these narratives come to become hegemonic. In particular, the article will challenge popular narratives of anything digital based on quasi-sublime and deterministic visions and cultures, advancing a focus on technologies' inner vulnerabilities, biases and material and design limitations. We will argue that

TECNOSCIENZA
Italian Journal of Science and Technology Studies
12 (2) pp. 79-87 - ISSN 2038-3460
www.tecnoscienza.net

T/S TECNOSCIENZA
Italian Journal of Science & Technology Studies

a stronger focus on technologies' weaknesses and vulnerabilities would bring beneficial insights to the current debates around digitization and its social and political impacts by bypassing polarized takes.

Since their birth, in fact, the rhetoric surrounding digital media and digital infrastructures has been constantly characterized by two main ideas. On the one hand, a long series of neologisms and metaphors have conveyed the idea of contemporary and future digital infrastructures as immaterial if not transcending worlds. Think for example at the early visions of the Internet as an "Intergalactic network" (Licklider 1963) or at the promises of the "cyberspace" (Mosco 2004). Or think of the recurring metaphor – harshly criticized, though still persistent - of "the cloud" (Peters 2015), up to the most recent techno-utopian dream of the "Metaverse" promoted by Mark Zuckerberg. Beside the political, ideological, or even metaphysical implications of this rhetoric (Natale and Pasulka 2019; Bory 2020), the social imaginary has been fiercely driven to think of infrastructures as distant, self-sufficient, and intangible means. On the other hand, companies, governments, and stakeholders have long characterized digital technologies through a series of adjectives and nouns evoking a sense of power, magnificence and reliability. Think at the unstoppable shift from mainframes to supercomputers or at the spread of the so-called information superhighways in the 1990s, the never-ending idea of a digital revolution, or the always imminent arrival of strong AIs (or again super-intelligence). Through these concepts, alterity and powerfulness, the term "digital" goes hand in hand with the idea of a distant and uncontrollable, but stable, efficient, and reliable system. As the sociologist Pierre Musso (2003) argues in his historical analysis of networking, this kind of rhetoric has both a fictional and functional implication. A system must be both narrated and perceived as strong and reliable. In other words, in our contemporary society, companies and providers need to instill trust in users and users need to blindly trust providers, otherwise any essential service would collapse or be replaced. Notably, infrastructures must be "off the radar, below notice, or off stage" (Peters 2015, 36); they must be strong, stable and reliable, allowing us to live our everyday lives with no concerns about the streets we walk on, the quality of the water we drink, and the data we access and share.

2. The Internet of Our Discontent and the Raise of Critical Takes

Today, after decades in which enthusiastic (if not ecstatic) visions of digitization have prevailed in the public sphere, critical scholars have challenged the propensity of digital technologies to strengthen individuals' protection and the democratic organization of societies. The Snowden case and the turmoil following the Cambridge Analytica (CA) controversy, for example, have inspired debates and discussions about Internet surveillance, the perils of the data economy and the potential "weaponization" of

social media platforms for political influence and propaganda goals. At the same time, the once supposed "horizontal" architecture of the Web has been clearly subverted by centralizing actors such as digital media corporations and national governments who exercise an immense power on our choices (Morozov 2011; Zuboff 2019). Moreover, the frequency in the use of the "black box" metaphor (Pasquale 2015) to define anything data or digital has increased significantly. Corporate algorithms are now defined as such, together with other controversial areas of datafication, such as surveillance, algorithmic manipulation, or machine learning. *Per se*, black boxes are socio-technical *apparata* capable of seeing and sensing all around them, without revealing enough information about their inner mechanisms. Not surprisingly, the metaphor works nicely when it comes to define how digital power is exercised by technological companies and other powerful actors, which are usually extremely successful in masking their actions, policies and dynamics behind veils of technical opaqueness and legal protections. As Ronald J. Deibert argues (2013, 5-9), never before have we known so little about how technology works, as we are actively discouraged from "developing a curiosity about and knowledge of the inner workings of cyberspace." Thus, it comes with little surprise that whistleblowers and leaks have taken a crucial public role in opening and exposing some of these black boxes, starting from Facebook (Olesen 2020). Overall, notwithstanding enthusiastic visions of the digital have been – at least in academia and in the media – completely overturned, the idea of digital media and infrastructures as strong and powerful has been rarely put into question[1].

Our aim here, as for the track we organized at the VIII STS Italia conference, is to look at digital media and infrastructures rather than through their strengths and power, through their vulnerabilities. From their side, STS have long been interested in the relationship between materiality and vulnerability, for example when addressing the relevance of repair and maintenance for the very existence of technical artifacts and infrastructures (Denis, Mongili and Pontille 2015; Russell and Vinsel 2018). More than a decade ago, in an article titled "The vulnerability of digital culture", Weibe Bijker already argued that "vulnerability is an inevitable characteristic of technological culture" but also that any vulnerability "is socially constructed as much as facts and artifacts are" (2006, 55-56). In line with Bijker's stance, we argue that understanding digital media and infrastructures as "weak" may help scholars to overcome the polarization of the goodness or evilness of technology. In our opinion, this peculiar perspective should start from analysing biases, bugs, and errors as essential elements of the systems we live by. Although some of these vulnerabilities appear in public discourse following incidents such as data breaches, outages, leaks, hacks, and other disruptive occurrences, sometimes they can also be the symptoms of more rooted phenomena and problems. For instance, the kind of problematic third-party data sharing that was at the core of the CA case was not an isolated incident, while actually a legitimate part of the Facebook business model at the time of the events. As many observers have

noted, the CA case was caused by a feature rather than a bug (AccessNow 2018). This point opens up interesting theoretical questions about almost two decades of hegemonic positivistic and deterministic takes on digitalization: have they been so "pervasive" to transform any discourse around digital things going wrong into an accident disrupting otherwise efficient and safe technologies and infrastructures?

3. Bubblegum and String: Infrastructural Spectacular Failures, Weaponization and Inherent Vice

Recent international media events have brought more attention on the vulnerabilities, bugs and errors of digitalization, shedding light on how these powerful systems can be weak and prone to malfunction. The global Facebook outage that occurred in October 2021 has definitely been one of the most interesting cases of this kind. As the web infrastructure and website security company Cloudflare wrote commenting the events (Martinho and Strickx 2021), seeing Facebook "disappearing from the Internet" has been probably the most explicit of these cases showing the existential weaknesses of today's digital infrastructures. As cybersecurity expert Eva Galperin noted, the accident also shown how "the internet is held together with bubblegum and string"[2], echoing recurring concerns about the stability and strengths of the Internet infrastructure. It is interesting to stress how the Facebook outage was caused completely by an internal mistake that occurred during a routine maintenance operation that disconnected Facebook data centers globally. In all its spectacularity, the biggest and richest global social network went completely offline by a rare but banal configuration mistake, underlining the hollowness of any "sublime" or "magical" view of digital infrastructures, the cloud or social networking at large. Other incidents had different origins. In 2016, for instance, the Mirai botnet brought interesting insights for a meta-analysis of the weaknesses of digitalization. Infrastructure company Dyn, offering DNS services to a set of major US clients, including Netflix, Amazon and PayPal among others, was targeted with a massive, distributed denial-of-service (DDoS) cyberattack, aimed at disrupting online services managed by Dyn. The result was a global outage that made enormous parts of the Internet unavailable for hours. While DDoS attacks are all but rare, this one was a peculiar one, as it was caused by a remotely controlled botnet of infected hijacked Internet of things (IoT) devices, such as printers, home appliances and security cameras (DeNardis 2020, 5-8). The malware Mirai was behind the infection of the devices involved in the botnet and it was created with the explicit aim of exploiting vulnerabilities in the devices' security, which is a topic of huge discussion in the field of IoT, given its usual low security standards (Bunz and Meikle 2018, 122). The Mirai botnet of *zombie* infected devices is so peculiar because it shows how inner digital vulnerabilities (i.e., weak

security standards) can be exploited remotely to launch attacks to the vulnerable Internet infrastructure.

4. "Mind" Vulnerabilities: Inside AI and Facial Recognition Shortsightedness

Vulnerabilities are not only a distinctive feature of digital infrastructures. If we adopt a simple and outdated analogy, the infrastructural "body" of digital systems is as weak as their "mind". For example, beyond the recurring myth of an upcoming superintelligence, humans' everyday life is constantly confronted with the biases and shortcomings of contemporary artificial intelligence such as voice assistants, facial recognition, social bots and companion robots. As scholars from different fields like anthropology, sociology, media and communication studies and STS have aptly shown, contemporary AIs often embeds the very same cultural biases and weaknesses of contemporary societies. Recent studies and critical enquiries have stressed how racism, deception, and western-centered behaviors and beliefs are among the many deficiencies of artificial intelligence, just like in our unequal and biased social world (Barassi 2020; Crawford, 2021; Crawford and Paglen 2021). This is clearly visible with facial recognition, one of the current most controversial applications of AIs, whose usage in various contexts has shown the existence of racial and gender biases in how the technology operates (Castelvecchi 2020). An influential study by Joy Buolamwini and Timnit Gebru (2018), for instance, underlined the presence of skin-type and gender biases in at least three commercial facial recognition systems. Similar results have emerged from further research and the available literature in this area is now extensive, as suggested by a comprehensive literature review by Khalil et al. (2020). Reasons for the presence of these persistent biases in facial recognition have to be found predominantly in the training materials that these systems are built upon and, in particular, in "internet-scraping at scale", the most frequently used approach to build large datasets for training facial recognition systems. These datasets, according to an Alan Turing Institute report, "have largely reflected the power relations, social hierarchies and differential structures of privilege that have together constituted the sociocultural reality from which those data were extracted in the first place" (Leslie 2020, 17-18). The profound ethical implications of biases in facial recognition, though, can also have severe civil rights consequences, especially when facial recognition is deployed as a law enforcement and security strategy in public spaces. In 2020, Robert Julian-Borchak Williams, a black man from the Detroit area, was wrongfully arrested after being falsely "recognized" by a facial recognition system in a CCTV footage (Hill 2020). The repressive and social sorting-oriented repercussions of facial recognition are even more explicit in China, where the technology has been used to target the oppressed Uighur minority. For instance, a 2020 *Washington Post*

investigation (Harwell and Dou 2020), based on internal documents, showed that a facial recognition software capable of sending automated "Uighur alarms" to the authorities had allegedly been tested in China. Yet, racial biases have emerged also in the application of other machine learning / AI applications, such as search algorithms: Safiya Umoja Noble's research work, among others, has demonstrated the existence of clear racist biases reinforcement and replicas in how commercial search engines like Google work, whose outcomes end up discriminating against minorities and black women in particular (Noble 2018, 64-110).

5. Conclusion: Joining Critical Voices to Unveil Digital Vulnerabilities

Especially in light of these profound ethical and societal concerns, deepening our understanding of the weaknesses and vulnerabilities of the "body" and the "mind" of contemporary digital systems means to reverse both the enthusiastic and the critical perspectives which indiscriminately accept the power of technology and its capacity of transcending human agency and social responsibility. Notably, by looking at weak systems scholars and policy makers can interrelate technological advancements and data infrastructures with human features and values to detect, acknowledge and even contain the very human errors embedded in contemporary sociotechnical systems. To think about the weakness and vulnerability of digital systems such as the Internet and AI is essential to understand how such systems, just like human societies, are quite far from reaching perfection, but they are, and they must, be mutually perfectible. On a broader societal level, though, a question about how technologies reach the public and how they get transformed into "narratives" remains unanswered. In particular, the role of media in perpetuating discourses and how they are created requires further scrutiny. For example, so far research conducted in the UK has shown how the public narrative of AI systems is predominantly driven by corporate and industrial interests and voices (Brennen, Howard and Nielsen 2018). Scholars and critical voices, by stressing and unveiling the vulnerabilities of weak systems, have the opportunity – and the duty – to influence and change these narratives. However, in order to counterbalance the overreaching voice of corporate actors, the inner weakness lying in academic fragmentation and disciplinary boundaries should be (respectfully) assessed and overcome. This article, which is a first result of the fruitful interdisciplinary panel we organized during the VIII STS Italia conference, is a first, short, step in such direction.

Acknowledgements

We are grateful to Veronica Barassi (University of St Gallen, Switzerland), Alex Dean Cybulski (University of Toronto), Louis Melançon (McGill University, Montreal), Stefania Milan (University of Amsterdam) and Maria Rikitianskaia (London School of Economics and Political Science) for joining our panel and for their great contribution and valuable inputs.

References

AccessNow (2018) *It's not a bug, it's a feature: How Cambridge Analytica demonstrates the desperate need for data protection*, in https://www.accessnow.org/its-not-a-bug-its-a-feature-how-cambridge-analytica-demonstrates-the-desperate-need-for-data-protection/ (retrieved December 30, 2021).

Barassi, V. (2020) *Child Data Citizen: How Tech Companies Are Profiling Us from Before Birth*, Cambridge, MIT Press.

Bijker, W.E. (2006) *The vulnerability of technological culture*, in H. Nowotny (ed.) *Cultures of Technology and the Quest for Innovation*, New York, Berghahn Books, pp. 52-69.

Bory, P. (2020) *The internet myth: From the internet imaginary to network ideologies*, London, University of Westminster Press.

Brennen, J.S., Howard, P.N. and Nielsen, R.K. (2018) *An industry-led debate: How uk media cover artificial intelligence*, in https://reutersinstitute.politics.ox.ac.uk/si-tes/default/files/2018-12/Brennen_UK_Media_Coverage_of_AI_FINAL.pdf (retrieved December 30, 2021).

Buolamwini, J. and Gebru, T. (2018) *Gender shades: Intersectional accuracy disparities in commercial gender classification*, in "Proceedings of the 1st Conference on Fairness, Accountability and Transparency", PMLR 81, pp. 77-91.

Bunz, M. and Meikle, G. (2018) *The Internet of Things*, Cambridge, Polity Press.

Castelvecchi, D. (2020) *Is facial recognition too biased to be let loose?,* in "Nature", 18 November, https://www.nature.com/arti-cles/d41586-020-03186-4 (retrieved December 30, 2021).

Crawford, K. (2021) *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven, Yale University Press.

Crawford, K. and Paglen, T. (2021) *Excavating AI: The politics of images in machine learning training sets*, in www.excavating.ai (retrieved December 30, 2021).

Deibert, R.J. (2013) *Black Code. Inside the Battle for Cyberspace*, Toronto, McClelland & Stewart.

DeNardis, L. (2020) *The internet in everything: Freedom and security in a world with no off switch*, New Haven, Yale University Press.

Denis, J., Mongili, A. and Pontille, D. (2016) *Maintenance & repair in science and technology studies*, in "Tecnoscienza: Italian Journal of Science & Technology Studies", 6 (2), pp. 5-16.

Harwell, D. and Dou, E. (2020) *Huawei tested AI software that could recognize Uighur minorities and alert police, report says*, in "The Washington Post", 8 December, https://www.washingtonpost.com/technology/2020/12/08/huawei-tested-ai-software-that-could-recognize-uighur-minorities-alert-police-report-says (retrieved December 30, 2021).

Hill, K. (2020) *Wrongfully accused by an algorithm*, in The New York Times, 24 June, https://www.nytimes.com/2020/06/24/tech-nology/facial-recognition-arrest.html (retrieved December 30, 2021).

Khalil, A., Ahmed, S.G., Khattak, A.M., and Al-Qirim, N. (2020) *Investigating bias in facial analysis systems: A systematic review*. In IEEE Access, 8, 130751-130761.

Leslie, D. (2020) *Understanding bias in facial recognition technologies: an explainer*, in "The Alan Turing Institute", https://doi.org/10.5281/zenodo. 4050457 (retrieved December 30, 2021).

Licklider, J.C.R. (1963) *Memorandum for: Members and affiliates of the intergalactic computer network; topics for discussion at the forthcoming meeting*, Washington, DC, Advanced Research Projects Agency, in https://www.kurzweilai.net/memorandum-for-members-and-affiliates-of-the-intergalactic-computer-network (retrieved December 30, 2021).

Magaudda, P. and Balbi, G. (2018) *Fallimenti digitali: un'archeologia dei "nuovi" media*, Milano, Unicopli.

Martinho, C. and Strickx, T. (2021) *Understanding how facebook disappeared from the internet*, in https://blog.cloudflare.com/october-2021-facebook-outage (retrieved December 30, 2021).

Morozov, E. (2011) *The Net Delusion: How Not to Liberate the World*, London, Penguin UK.

Mosco, V. (2004) *The Digital Sublime. Myth, Power, and Cyberspace*, Cambridge, MA, MIT Press.

Musso, P. (2003) *Critique des réseaux*, Paris, Presses Universitaires de France.

Natale, S. and Pasulka, D. (Eds.) (2019) *Believing in Bits: Digital Media and the Supernatural*, Oxford, Oxford University Press.

Noble, S. U. (2018) *Algorithms of Oppression. How Search Engines Reinforce Racism*, New York, New York University Press.

Olesen, T. (2020) *Whistleblowing in a time of digital (in)visibility: towards a sociology of 'grey areas'*, in "Information, Communication & Society", Online first, https://doi.org/10.1080/1369118X.2020.1787484.

Pasquale, F. (2015) *The Black Box Society. The Secret Algorithms that Control Money and Information*, Cambridge, MA, Harvard University Press.

Peters, J. D. (2015) *The Marvelous Cloud*, Chicago, University of Chicago Press.

Russell, A. L. and Vinsel, L. (2018) *After Innovation, Turn to Maintenance*, in "Technology and Culture", 59 (1), pp. 1-25.

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York, Profile books.

[1] Scholars from STS and media history have partially filled this gap by looking symmetrically at the relevance of failures and forgotten projects to the development of the contemporary socio-technical environment (see Magaudda and Balbi 2018).

[2] Eva Galperin's tweet is available at: https://twitter.com/evacide/status/1445177126139809796 (retrieved December 30, 2021).