Erosion of journalistic standards

In my homeland of Israel, three bribery indictments against Prime Minister Netanyahu have sent the political system into turmoil, resulting in three elections over the course of twelve months. During this year, I was overwhelmed by the amount and intensity of social media posts accusing politicians and government officials of wrongdoings. These posts often contradicted one another, were unreferenced and originated from anonymous sources.

Attempts to validate the content of these posts with credible sources did not always succeed, and I was left to wonder whether they are true or false.

While this example relates to Israeli politics, disinformation, defined here as incorrect or biased information disseminated unintentionally or intentionally, is far more ubiquitous and affects additional aspects of life on a global scale, particularly the reputations of brands and firms [1], and public health [2].

Communication and expression through various written media without the involvement of the government is a human right protected under the freedom of the press [3]. The purpose of this freedom is clear: free written communication allows the exchange of information, debate, and critique that are essential to scientific and social advances as well as a functioning opposition in democratic systems. In its current form, however, freedom of the press achieves these goals only if the press is subject to strict standards of veracity and balance. While these standards are usually maintained by self-regulation in traditional print and broadcast, they are not always followed by social media users, and in some cases, are deliberately violated. As a consequence, information on social media is littered with disinformation, which contributes to the polarization of public views and opinions and promotes extremism [4].

In this essay, I will argue in defense of the freedom of the press in the age of social media. To defend it, three censorship-free strategies, focused on regulation of the context rather than the content of information are suggested to mitigate the spreading of disinformation and its subsequent negative consequences.

Social media catalyzes disinformation propagation

Prior to social media, dissemination of information to the public was done mainly through print and broadcast, for which the entry
barriers for new players are high. Traditional media companies were relatively few and known within their communities, and reputation had significant role in their marketing campaigns. This has incentivized the media companies to adhere to strict journalistic standards to prevent accidental dissemination of disinformation, which would have stained their reputation.

Social media facilitates an effective, multilateral mode of communication by enabling any individual or organization to post, comment or share information. This mode of communication has numerous advantages for individuals, groups, and firms [5]. However, it also lowers the entry barriers for disseminating information, thereby eroding journalistic standards and, subsequently, the reliability of information in social media.

The ubiquity of unreliable information in social media is greatly problematic because many of us get at least some of our news from social media [6], and often fail to distinguish between reliable and unreliable sources of information [7]. Furthermore, a number of sociological and technological factors aggravate the problem of disinformation on social media at the societal level. First, we tend to search for, uncritically accept, and recall information that is aligned with our views and beliefs, even if it is wrong, more often than information that is not aligned with our views and beliefs, a phenomenon known as confirmation bias [8]. Second, we tend to focus on a limited number of sources of information and interact with people who rely on the same sources [9]. This creates “echo chambers” where the views and beliefs of individuals are constantly reinforced and amplified by their peers, which promotes polarization and extremism, and decreases trust in other users and sources of information [10]. Third, when malicious intents are taken into account, propagation of disinformation can be augmented by using social bots (fake accounts operated by artificial intelligence) to trick social media algorithms into expanding the exposure of disinformation. This increases the chance of disinformation going viral and being propagated by innocent users, which consequently extends its exposure by orders of magnitude [11]. Approximately 5% of Facebook accounts [12], and between 9-15% of Twitter accounts [13], are fake, a testament to the severity of the problem.

Due to our inability to correctly evaluate information accuracy, along with the socio-technological factors that promote exposure to disinformation, social media serves as a catalyst for the propagation of disinformation. Luckily, this can be changed by adopting the policies suggested below.

**Strategies for mitigating the societal effects of disinformation**

**Strategy 1 - Incentivize self-regulation of social media platforms**

Disinformation existed well before the internet as a method for damaging the reputation of individuals, groups, and firms [14]. As a countermeasure, anti-defamation laws were developed to allow victims to claim compensation from defamers. However, these laws are not suited for the digital age, in which social media provides the technical means to disseminate disinformation anonymously, rendering legal actions against defamers futile.

Furthermore, US-based social media companies are not liable for defamatory posts of users on their platforms [15]. Nowadays, incentivizing self-regulation of social media posts by reforming anti-defamation laws so that victims of viral disinformation can sue the platforms, may be useful. This will motivate social media companies to collaborate with researchers and seek automated solutions to self-regulate their platforms and prevent the spreading of disinformation, at least of defamatory nature.

Social media platforms track user activity in great detail and invest immense resources in analyzing this data to produce personalized advertisements. Many academic studies have found that user activity data can be used to automatically detect many forms of disinformation [16]; however, social media companies do not currently use this data to prevent spreading of disinformation. In fact, Facebook’s Community Standards [17] and Tweeter Rules [18] do not prohibit general
disinformation, demonstrating that this is not a priority of the companies. Furthermore, viral disinformation is engaging and may prompt users to spend more time on the platform, thus incentivizing social media companies against taking a stand to discourage disinformation. Reforming the incentive structure of social media companies will instead urge them to prevent disinformation propagation.

**Strategy 2 – Coupling viral posts with fact-checking reports**

The ubiquity of false claims shared on social media platforms has motivated individuals and organizations such as bloggers, dedicated websites, and newspapers to fact-check viral content and publish publicly-accessible, detailed reports on these posts. On the surface, such fact-checking reports can mitigate the damage caused by disinformation. In practice, however, they are not effective. We are not likely to share fact-checking reports if they do not support our views [19]. Conversely, due to confirmation bias, we are likely to uncritically accept (without consulting fact-checking reports) and share disinformation if it aligns with our views and beliefs [20]. Disinformation and follow-up fact-checking reports are thus spread to different communities; disinformation to communities that uncritically accept it and its refutation to communities that criticize the disinformation and hinders its propagation.

A possible solution to this problem is to couple the propagation of disinformation and its refute by tagging posts according to their nature (e.g. facts, opinions, satire), accuracy, and objectivity by certified taggers. The tagged posts should also be linked to the full informative fact-checking reports to provide users with further details. Of course, tagging should not be a prerequisite for posting, which can be tag-free; however, once a post has been tagged, the tags should appear on all of its shared copies. Such tags serve two purposes: first, tagged disinformation will be less likely to be accepted and shared; second, credible information that refutes disinformation will be directly linked to the false posts, thereby facilitating transfer of information across communities, and reducing echo-chamber effects.

Implementation of this mechanism is expected to face two major obstacles. First, social media companies are likely to be uncooperative if these tags reduce user engagement. In that case, cooperation can be incentivized using strategy 1 above, or be required by regulation. Second, the identity of the individuals and organizations certified to tag posts is a major concern. Studies show that professional journalists [21] and fact-checking services [22] are not collectively trusted by the public. Therefore, certifying them would result in loss of trust by skeptical users, who may simply ignore the tags, rendering this suggestion ineffective. Therefore, a freely-accessible, crowd-sourced solution that minimizes bias is preferred. Wikipedia-like collaborative model, in which the public can freely contribute to preparation of fact-checking reports that refer to social media posts, may be effective. However, to avoid trolling and mistakes, only certified taggers democratically elected by contributing members will have the authority to tag posts on social media and link them to reports. Wikipedia’s status as the most popular encyclopedia attests to the public’s trust in such collaborative work. Furthermore, academic studies of Wikipedia show that it can produce high-accuracy content [23] quickly (especially news-related content) [24], and effectively self-regulate disinformation regarding matters of public interest [25]. Hopefully, the public trust and performance of Wikipedia can be replicated in fact-checking contexts.

**Strategy 3 – promote digital media literacy**

Due to political and bureaucratic reasons, school curricula around the world do not prepare students for the challenges of the 21st century. In Israel for example, the curriculum includes religious studies and literature as core subjects, while science and technology-oriented classes are elective. In order to equip students with the necessary skills for responsibly using social media as an information source, concepts such as sponsored advertisement, cognitive biases, and fact-checking must be taught. Recent study shows that introduction of such material
into the curriculum indeed improves students’ ability to critically evaluate online information [26]. Therefore, nationwide adoption of appropriate school curriculum is a long-term instrument to mitigate negative outcomes of disinformation, which will become increasingly effective as more generations of students graduate. The main challenge in nationwide addition of a new subject to the curriculum is training and hiring able teachers, which is a slow and costly process. To overcome this difficulty, the course can be developed by relevant academic researchers as an obligatory, online course similar to those offered by Coursera. That way, not only can the course be introduced quickly and at low costs, it can also be updated regularly to include future emerging topics. Additionally, making the course publicly-accessible introduces added value for curious adults at minimal costs. In this way, a growing portion of society will be able to critically evaluate information.

Concluding remarks

If there is one thing that history has taught us, it is that inventions are only as good or bad as the people who use them. Alfred Nobel’s invention of the dynamite has revolutionized construction and mining, but also warfare. Similarly, social media has assisted civil uprising against dictatorships in the Arab spring, and enabled women to group and oppose sexually abusive environments worldwide via the #MeToo campaign [27]. On the other hand, social media is polarizing and is partially to blame for the populist wave that washed over global politics [28], and the growing popularity of the anti-vaccination movement [2], which have led to widespread political unrest and the resurgence of almost eradicated diseases. The key to the effect of social media on the future lies in the policies we, as a society, will adopt.

This essay suggested three complementary strategies to mitigate the negative effects of disinformation in social media. First, reform of anti-defamation laws will incentivize social media companies to develop automatic solutions for detection and blocking of defamatory disinformation. This can be accomplished in the short term, but it is a partial solution since it only refers to libel. Second, development of a publicly trusted fact-checking mechanism to evaluate and tag viral posts is a powerful mechanism to oppose disinformation propagation and echo-chambers. However, it relies on the cooperation of social media companies which may need to be properly incentivized or regulated, and on the development of such fact-checking mechanism. This is expected to take several years at the very least, making this strategy an intermediate-term one. Finally, we suggest to modify school curricula and adapt them to the social media era, with online digital literacy programs. Education is possibly the most effective tool to immunize society from being affected by disinformation. However, it mainly affects students, making it a long-term strategy that must rely on the time bought by the first two strategies.

Hopefully, adopting these three strategies will direct us toward a brighter future, in which social media improves communication, but does not lead to propagation of disinformation and polarization of public discourse.
Bibliography


