Safeguarding Information Integrity:
Addressing Externality Costs of Misinformation in the Digital Age

Introduction

The fabric of our global society is interwoven with the threads of information, and its integrity is paramount to the functioning of democratic institutions and the preservation of public trust. With the advent of social media, the ease, speed and reach of dis/misinformation dissemination has become a pervasive force, challenging the United Nations' commitment to peace, security, and responsible communication. In this submission, ICT4Peace considers the ad-revenue models of social media platforms that amplify misinformation and generate externality costs and proposes innovative solutions to mitigate these effects.

The Impact on Democratic Processes and Public Trust

The digital age has ushered in transformative shifts in how information is disseminated and consumed, significantly impacting democratic processes and public trust. A study in Humanities and Social Sciences Communications Journal found that misinformation is typically characterized by reduced cognitive complexity and heightened emotional arousal, making it more digestible and shareable among users. This ease of consumption and emotional resonance amplifies misinformation's reach, compromising the quality of public discourse and eroding trust in democratic institutions.

Externalities of Dis/Misinformation

Like industrial pollutants that harm the environment, the spread of misinformation on social media inflicts societal and human security costs. Social media platforms effectively create information systems that not only generate revenue from the proliferation of misinformation but does so without bearing the societal costs of its spread. The analogy to environmental externalities is apt; just as industries may emit pollutants into the environment without bearing the full cost of the resultant damage, social media platforms benefit financially from the engagement generated by misinformation without accounting for its societal impact. This lack of accountability has been highlighted in the context of climate dis/misinformation (so generating double-externalities) by organizations such as Friends of the Earth, Avaaz, and Greenpeace USA, who point to the gross lack of transparency and the inadequate measures taken by social media platforms to combat the spread of such damaging narratives. The need for robust standards is evident, as is the need for transparency from social media companies, which is key to understanding the evolving landscape of disinformation and holding disseminators accountable

Externalities of Dis/Misinformation on X

Social media platform X exemplifies the challenges facing digital information ecosystems. Following a change in leadership and the subsequent dismantling of misinformation safeguards, X has seen an uptick in the spread of misleading content. The transformation of the blue check mark system has been instrumental in this shift, where anyone can pay to have their posts algorithmically prioritized, regardless of content accuracy, and even benefit financially from them.

The inclusion of the “Ads Revenue Sharing feature” in X's paid Premium and Premium+ tiers incentivizes users to create content that maximizes engagement and clicks, as individual users have the possibility to directly benefit financially. A striking illustration of this is the dissemination of false narratives during the Israel-Hamas conflict, where a significant majority of the most viral misinformation posts were traced back to paid subscribers. Furthermore, the Center for Countering Digital Hate reported a spike in tweets containing slurs, indicating a broader trend of harmful content proliferation under the new revenue-focused policies.

In this context, X's financial model mirrors the issue of environmental externalities, where companies benefit from activities that impose significant costs on society—such as pollution or social division—without paying for the damage done. This analogy underscores the urgent need for regulatory frameworks to address these 'digital externalities', compelling platforms to assume responsibility for the misinformation disseminated through their networks.

Proposed Responses and Recommendations

The pervasive nature of misinformation necessitates a robust and multifaceted response strategy. To this end, ICT4Peace makes the following recommendations.

Misinformation Impact Assessment Model

To address the societal costs of misinformation, we propose a model that imposes a cost on digital platforms based on the prevalence of misinformation they disseminate. This model consists of a 'Misinformation Impact Assessment' to measure the extent and impact of misinformation on each platform. The model would not only quantify the prevalence of misinformation but also its societal repercussions, akin to an environmental impact assessment that measures the effects of industrial activities on ecosystems. Platforms would then make a financial contribution proportional to the severity of misinformation detected. The funds collected would be allocated to initiatives combatting misinformation, such as supporting reporting by independent journalism, oversight by independent CSOs, funding for digital literacy programs, and advancing misinformation research.

As part of their adherence to the proposed Code of Conduct, online platforms would agree to participate in the program, and contribute financially in the initial stage to support the development of the model. This approach not only incentivizes platforms to more effectively manage misinformation but also ensures they contribute to offsetting its broader societal impacts. Regular audits and adjustments would be conducted to ensure accurate assessments and to adapt to the evolving nature of online misinformation. The process would be transparent, with clear accountability mechanisms overseeing the assessment and fund allocation. The development and oversight of the model would be managed by one or more CSOs with relevant expertise and a respected human-rights promoted reputation, and that is independent of funding from ICT companies or governments that routinely generate or spread dis/misinformation in order to avoid compromising conflicts of interest.

Advanced Natural Language Processing

Another tool in the information integrity tool box would be to use advanced Natural Language Processing (NLP) techniques, as detailed in the Humanities and Social Sciences Communications Journal article. These NLP techniques could proactively identify potential misinformation by analyzing readability, complexity, and emotional content. This preemptive approach would allow platforms to curb the virality of misinformation before it spreads, aligning with the public interest of maintaining a fact-based information environment.

Information Integrity Capacity-Building

Furthermore, capacity-building programs are essential in cultivating a discerning online populace. By integrating digital literacy and critical thinking curricula in schools and incentivizing employers to require staff to enroll in professional development programs, individuals can be better equipped to navigate the information landscape. This educational initiative would empower users to critically evaluate content, understand the nuances of misinformation, and engage with information more responsibly. Funding for this could come in part from the above-mentioned Misinformation Impact Assessment program.

Inclusion of Dis/Misinformation in Discussions of relevant UN Initiatives

Finally, we recommend that dis/misinformation be explicitly included in the discussions of relevant UN initiatives. At the substantive meeting of the Open Ended Working Group on ICTs in July, Ambassador Gafoor did a masterful job of achieving consensus on the second Annual Progress Report (APR). However, the threat of misinformation, previously acknowledged in the Zero Drafts of the 2nd APR, was substituted with a narrower focus on State-led "information campaigns" in the APR that was adopted. This omission overlooks the multifaceted nature of mis/disinformation threats, which extend beyond state actions and permeate non-state action and impact and public discourse. Furthermore, the omission minimizes the extremely important threat of mis and disinformation in a multitude of areas, from national elections to human rights protection to armed conflict, which we have seen extensively in the Ukraine conflict, and now more recently in the conflict between Israel and Hamas. Particularly regarding this last point, we also recommend that the UN Security Council take up this matter, as dis/misinformation can have a devastating impact on international security.

Conclusion

In conclusion, this submission has underscored the pressing issue of misinformation in the digital age, likening its societal impacts to environmental externalities that go unchecked. The social media platform X's case study illustrates the urgent need for responsible platform governance and the recalibration of reward systems towards truth and accuracy. The 'Misinformation Impact Assessment' model, bolstered by Natural Language Processing, offers a promising avenue for prevention, early detection and quantification of misinformation. Complementing technological solutions with capacity-building educational programs can foster an informed citizenry equipped to engage with information critically, thereby lessening the negative impacts of dis/misinformation. Finally, the risks and threats of dis/misinformation need to be systematically included in relevant discussions within the UN. Collectively, these efforts constitute a proactive step towards a digital ecosystem where responsible dissemination of information is better supported, and platforms are held accountable for upholding information integrity.

Anne-Marie Buzatu
Executive Director, ICT4Peace Foundation