
DEFAMATION BY ALGORITHM: LIABILITY OF PLATFORMS FOR AUTOMATED CONTENT AMPLIFICATION

Tanisha Acharya, B.B.A. LL.B. (Hons), Symbiosis Law School, Noida

ABSTRACT

The recent developments which have taken place in the digital communication platforms have shifted the role of platforms from passive intermediaries to active distributors of information, fundamentally altering the public discourse. At the core of this transformation lies a set of algorithmic systems that amplify and curate content strategically to drive user engagement. Through this paper, we will analyse the problem caused to the general public from “defamation by algorithm”, and we will also examine if the ambit of the term “publication” extends to the automatic amplification of libelous content and the liability of platforms for dissemination of such content. The priority given by these data driven systems to viral metrics rather than focusing on factual relevance of the content leads to character assassination spreading like wildfire across borders. This has the potential of causing irreversible reputational damage, rendering the judicial remedies obsolete. The adequacy of the current legal frameworks in maintaining a balance between right to reputation and freedom of speech and expression is evaluated through this paper by adopting a doctrinal and analytical methodology. This paper suggests that even if the lack of human intent is used as a defence by the platforms, similarity can be found between traditional editorial publication and the targeted promotion of content by various algorithms. The study finds that the opacity of algorithmic harms can’t be addressed by traditional defamation principles and “actual knowledge” standards as they are not well equipped to do so. As a result, it has become a struggle for the courts all around the world to hold algorithmic systems accountable within outdated statutory parameters. The paper highlights the need for a rights-based regulatory model which can balance accountability of the platforms without compromising Article-19 which talks about freedom of speech and expression and legal reforms which can recognize the role of algorithms in disseminating content.

Keywords: Algorithmic Amplification, Defamation, Intermediary Liability, Safe Harbor, Digital Services Act, Publication.

Introduction

In today's world, digital platforms make widespread use of algorithmic systems to curate the content, rank them and amplify it. It is the work of these automated systems to decide what information will reach the audiences. On one hand, while algorithms maximise efficiency, they also generate libelous content which harms an individual or organisation's reputation.

The case of traditional defamation, where human authors or editors were held accountable, is quite different from the defamation caused by algorithms as it poses serious legal questions. Historically, defamation was limited only to human agency where composition could be made only by writer's decision or by the approval of editor to publish a particular statement. However, modern technology has upgraded the level of defamation. A common defence given by platforms in such cases is that they lack intent and that the information is deployed automatically, serving just as a passive channel for the information to flow through. Nonetheless, the impact such defamation has is still severe as within a matter of few minutes, they could promote deceptive content which could cause widespread harm to a company's reputation. The modern algorithm is very different from traditional media as the circulation of content during the time when traditional media was prevalent took place through physical distribution networks but modern algorithms on the other hand, prioritize sensational and highly provocative content which can maximise user engagement. By doing so, the chances of transmitting unverified, defamatory content also increases. The problem arises when it becomes difficult for the human moderators to regulate this content which poses the danger of becoming a globally accessible piece of disinformation.

Furthermore, the advancement in technology has caused the platform to become an active intermediary from a passive host. This is because the algorithm is not just holding data anymore; but it is also involved in its distribution to targeted audiences through personalized feeds. This is done by pushing the defamatory matter into their viewer's feed to maximize user engagement with the content and expand its reach platform-wide.

This paper investigates how traditional defamation principles apply to the algorithmic systems, it examines if the platform is liable for the defamatory content and if current legal principles are capable enough to handle this new issue.

Issue

1. Whether the algorithmic system's amplification of content will constitute "publication" under defamation law.
2. Can digital platforms be held accountable for defamatory content when automated algorithms disseminate the content without human involvement?
3. Can the lack of malice or intent in algorithmic systems exempt the platforms from liability for defamation?
4. Can platforms assert intermediary safe harbour protection when the generated defamatory content is clearly amplified by their algorithm?
5. Are the existing defamation liability laws sufficient to tackle the harms arising out of automated content amplification?

Rule

Defamation- It is the publication of a false statement which is made without the consent of the person(s) defamed and which tends to lower the reputation of the person(s) in the minds of the right-thinking members of the society. Even if the certain person(s) is not named but it is ascertainable as to whom the defamatory statement is intended towards, it will still count as defamation. Key Essentials of Defamation –

- The statement must be defamatory and false. If the statement is a truth, it won't be considered as defamation.
- The defamatory statement must clearly be intended towards the plaintiff even if the name of the plaintiff is not mentioned.
- The statement must be published, i.e. it should be made in front of a third party.

"Section 79 of the Information Technology Act"- Intermediaries that inculcate due diligence are granted safe harbour protection under Section 79 of the Information Technology Act, 2000. In the case of *Shreya Singhal V. Union of India*, the supreme Court upheld these safeguards conditioned on actual knowledge of illegal content and timely action upon notice.

“Section 230 of the Communications Decency Act, 1996”- In the US, the platforms are shielded from user generated content by the help of Section 230 of the Communications Decency Act, 1996.

“Digital Services Act”- In the European Union, the European Union: The Digital Services Act ensures that large platforms take adequate precautions to reduce systemic risks caused from content dissemination.

“McAlpine v Bercow”- The court ruled that even in scenarios where there has been an indirect communication, there will be publication of the defamatory content posted on social media.

“Shreya Singhal V. Union of India”- The action of removing content following a personal takedown request is not necessary for online intermediaries. However, the intermediaries will be held accountable if even after the receiving the court’s order or a government agency’s order, they have failed to take down the defamatory content.

“Delfi AS v Estonia”- A platform was held liable in the case of Delfi AS v. Estonia by the European Court of Human Rights for its failure in regulating defamatory user remarks and avoiding foreseeable harm.

Application

The first question with regard to defamation law, is whether algorithmic amplification classifies as defamation. According to traditional defamation, it was necessary for a third party to be informed about the defamatory statement but in today’s world of technology, courts have made it clear that for defamation to occur, there is no need for direct human communication to constitute publication. As established in the case of **“McAlpine V. Bercow¹”**, even indirect communication that takes place via social media platforms can constitute publication. The curation, ranking, and content recommendation by algorithms, increases the visibility and reach of the content. As a result, amplification of content by the algorithms can contribute to publication as the platforms actively help in spreading the defamatory message to a larger audience.

The second question is can digital platforms be held liable when there is no human involvement

¹ McAlpine v. Bercow, [2013] EWHC 1342 (QB).

involved in disseminating the defamatory content. We must understand one thing that platforms are responsible for managing algorithms even though they operate on their own. The increased engagement generated by the amplification doesn't necessarily negate liability due to the lack of human intervention because for defamation to occur, it is not always necessary for intent to be there. The failure to prevent foreseeable harm or negligence can also result in liability. Therefore, platforms cannot take the defence of automated nature of algorithm to absolve themselves from liability because if there is a harm to the reputation of a person, then it doesn't matter if the dissemination of the content was intentional or automated.

The third question is can the lack of malice or intent in algorithmic systems exempt the platforms from liability for defamation. For answering this we must refer to "section 79 of the Information Technology Act" which mentions that intermediaries who exhibit due diligence with respect to intermediary safe harbour protection are granted immunity. In the case of "**Shreya Singhal V. Union of India**"², the Supreme Court made it clear that it is not necessary for middlemen to comply with private takedown requests. It is only when the intermediaries defy a court order or a communication from an authorised government agency and fail to take down unlawful content that they become liable. Hence, the safe harbour protection is given only until the defamatory content continues to be algorithmically amplified even after receiving a legal notice.

Comparatively, in the United States, broader immunity is provided to the platforms for user generated content where even the algorithms are involved, under "Section 230 of the Communications Decency Act, 1996" which is unlike the European approach which is stricter. The European Court of Human Rights held a platform liable in the case of "**Delfi AS v. Estonia**"³ for not regulating offensive user remarks and preventing predictable harm. The Digital Services Act further strengthens this approach by compelling big platforms to reduce systemic risks caused due to automated content generation.

Finally, we must understand that in today's generation, it becomes really necessary to assess the effectiveness of current defamation laws. The speed, scope and opacity of automated systems were not meant to be dealt by traditional defamation standards, even though they are somewhat applicable on algorithmic amplification. The contrast in strategies adopted by India,

² Shreya Singhal v. Union of India, (2015) 5 SCC 1.

³ Delfi AS v. Estonia, App. No. 64569/09, Eur. Ct. H.R. (2015).

US and EU shows the difficulty of current institutions in striking a balance between accountability, free speech and platform immunity.

Conclusion

The traditional legal principles are greatly challenged due to defamation caused by algorithm. Despite, there being lack of intent on the part of the algorithms, their role in spreading and amplifying content spreads the idea that platforms could be viewed as publishers for the purpose of defamation law. Safe harbour protections like the ones included under “Sec 230 of the Communications Decency Act” and “Section 79 of the Information Technology Act” are vital in safeguarding the intermediaries, but they are not effective always because in certain conditions the platforms continue to spread false information even after they are given a legal notice.

The European strategy reflects a shift towards more platform accountability in dealing with the systemic damages generate by automated content dissemination, as established in the case of “Delfi AS v Estonia” and the “Digital services Act.” In comparison to this, the framework of the US and Indian Systems offer better protection but lack in dealing with defamation caused by the algorithm. Hence, the current defamation laws are not well prepared to handle the issues caused by automated content amplification even after providing a basic framework. This calls in for legal reforms that understand the role of algorithms and hold platforms accountable while recognising the importance of freedom of speech.

Bibliography

Books

Ratanlal & Dhiraj Lal, *The Law of Torts* (LexisNexis, latest ed.).

W.V.H. Rogers, *Winfield and Jolowicz on Tort* (Sweet & Maxwell, latest ed.).

M.P. Jain, *Indian Constitutional Law* (LexisNexis, latest ed.).

Statutes and Legal Instruments

Information Technology Act, No. 21 of 2000, § 79 (India).

Communications Decency Act, 47 U.S.C. § 230 (1996).

Regulation (EU) 2022/2065 of the European Parliament and of the Council (Digital Services Act).

Case Laws

Shreya Singhal v. Union of India, (2015) 5 SCC 1 (India).

McAlpine v. Bercow, [2013] EWHC 1342 (QB).

Delfi AS v. Estonia, App. No. 64569/09, Eur. Ct. H.R. (2015).

Websites

European Commission, *Digital Services Act*, <https://digital-strategy.ec.europa.eu>.

Electronic Frontier Foundation, *Section 230 Explained*, <https://www.eff.org>.

Ministry of Electronics & Information Technology, Government of India, <https://www.meity.gov.in>