

Review of: "Exploding AI-Generated Deepfakes and Misinformation: A Threat to Global Concern in the 21st Century"

Maria Teresa Andrade¹

¹ Universidade do Porto

Potential competing interests: No potential competing interests to declare.

Generally, the level of writing is very good, and the addressed topic is very relevant and interesting. But, in my opinion, the article has limited added value because it does not provide meaningful insights into research being conducted and results obtained in detecting and preventing diverse types of deep fakes. It is a good informative source of high-level concepts and approaches dealing with deep fakes but is meant for a non-expert audience. Many of the provided references are not appropriate, and in the article, the author does not provide a meaningful analysis of the work and results that have been presented in some of the interesting referenced articles. It thus has limited value for researchers and technologists.

Specific comments on a chapter basis follow below.

In chapter 2:

- Reference 5, when addressing GANs, is not adequate. And although the description of GANs is well-crafted and provides a concise overview of the key concepts, it is very high level and does not provide insights on how GANs are actually used for deep fakes.
- There is no reference for Deep Neural Networks.
- Reference 6 is not appropriate.

In chapter 3:

- Apart from reference 3, no other reference is given in this section, not even on the very important aspect of the existing challenges in detecting deep fakes and counteracting them, which was stated in the abstract as an objective of the study conducted

It is not clear the reason for having chapter 4.

In chapter 5:

Generic descriptions of the various types of deepfakes are given, with only 2 references, both text-based, which are additionally inappropriate. The added value is minimal.

Chapter 6, addressing the impacts, does not provide additional information in relation to the previous chapters.

In chapter 7:

Reference 1 is appropriate. However, almost nothing is described concerning the methods, work, and results described in that article. No other references are provided, and only a brief introduction is presented concerning several possible approaches to detect and mitigate deep fakes. The added value is minimal as the reader is not informed of the validity and performance of such approaches nor the operational details.

In chapter 8:

References are inappropriate.

Chapter 10:

Very limited added value as the recommendations are very high-level. One good reference is given (ref 15), but there is no real analysis of the work and results described in that article.

The section “Do’s and Don’ts” is interesting but again presents quite common sense ideas.