



INCLUSIVE SOCIETY INSTITUTE

Op-ed

AI AND SOCIAL MEDIA: A DOUBLE-EDGED SWORD THAT DEMANDS GLOBAL RESPONSIBILITY

By Daryl Swanepoel

Artificial Intelligence (AI) is rapidly transforming social media. It powers recommendation engines, moderates content, personalises feeds and generates posts so convincingly that the line between real and fake is now dangerously thin. But while this technology can democratise access to information and connect people more deeply, it is increasingly becoming a threat to truth, social cohesion and even democratic stability.

AI in social media is both a force for good and a force for harm. It is here to stay, but without decisive regulation and oversight, the damage it inflicts may outpace its benefits.

At its best, AI enhances accessibility, assists content moderation and improves user experience. It can help detect hate speech, flag disinformation and promote social campaigns that might otherwise remain unseen. It allows for real-time translation, speech recognition and tailored information flows that can make the digital space more inclusive.

But at its worst, it is a machine that learns to exploit outrage. Social media algorithms - now heavily AI-driven - prioritise engagement over truth. The more emotionally charged a post is, the more traction it gains. Add to this the rise of deepfakes, voice cloning and AI-written texts and we find ourselves in a post-truth environment where lies spread faster than fact. Often, these lies are only exposed after the damage has already been done.

This danger is not theoretical. In South Africa, we've seen how inflammatory social media content, amplified by algorithmic echo chambers, contributed to the July 2021 riots, where more than 300 lives were lost, and thousands of businesses destroyed, yet many of the mobilising posts were later found to have been misleading or entirely fabricated. Similarly, xenophobic violence has been fuelled by viral posts accusing migrants of crimes or "stealing jobs," usually with little evidence.

The issues of land reform, employment equity and farm attacks have likewise been distorted through online misinformation campaigns. AI does not distinguish between socially responsible content and divisive propaganda. It simply amplifies what is most clickable.

In response, there is an urgent need to mandate algorithmic transparency and pluralism and platforms should be required to present multiple sides of complex issues, not just the one that reinforces a user's existing worldview. This isn't about censorship; it's about creating informed citizens.

Furthermore, regulators must insist that users are notified when they are interacting with AI-generated content. AI systems that amplify or generate political content should be subject to strict transparency standards, because without this, disinformation will continue to masquerade as grassroots opinion, through which democratic accountability will be further eroded.

At a national level, countries like South Africa must begin treating AI governance as a core public policy priority. But even more importantly, this cannot be left to national governments alone.

The role of the G20 and global governance

Given the global nature of digital platforms and the borderless spread of misinformation, this is a matter of international urgency. The G20, an institution which brings together the world's largest economies, is uniquely positioned to play a leading role in establishing a '*Global Framework on Ethical AI and Algorithmic Accountability*'.

At the 2024 G20 Summit, several member states signalled concern over AI's risks, but concrete regulatory coordination remains weak and this must change. The G20 should endorse principles requiring:

- Algorithmic explainability: Users have the right to know why certain content is shown to them.
- Pluralism by design: Algorithms must be required to include counter-narratives on divisive issues.
- AI origin tags: AI-generated content must be clearly labelled.
- Cross-border cooperation: Nations must collaborate on AI misuse, especially in electoral contexts.

Without shared governance standards, authoritarian regimes and malicious actors will continue to exploit open platforms to manipulate public opinion in democracies. Multilateral cooperation is the only sustainable path forward.

A South African lens

South Africa's own fragile social fabric makes it especially vulnerable to AI-driven misinformation. The erosion of trust in public institutions, highlighted in the 2025 South African Social Cohesion Index (SASCI), is exacerbated by a social media landscape where falsehoods too often dominate the discourse.

To address this, South Africa should:

- Integrate AI literacy into school curricula and public campaigns.
- Establish a Digital Ethics Commission to oversee responsible AI use in media and political communication.
- Require platforms operating in the country to disclose AI involvement in content shaping.
- Advocate at the UN and G20 for binding international AI standards that protect democratic processes.

Civil society and independent media also have a crucial role to play. Fact-checking organisations must be supported and scaled. Media literacy should be treated as essential civic education. And the public must be encouraged to value verification over virality.

Conclusion

AI's integration into social media is not optional. It is already happening and accelerating. But whether it builds stronger societies or more polarised ones depends entirely on how we regulate it.

We need a global conversation, rooted in ethical responsibility, democratic values, and practical safeguards. The G20, and countries like South Africa, must lead this dialogue.

Left to its own devices, AI will continue to serve the logic of engagement over the imperatives of truth, but if we act collectively, we can harness it for the public good.

The stakes are too high for complacency. In the age of AI, our democracy, our social cohesion and even our sense of shared reality hang in the balance.

Daryl Swanepoel is the Chief Executive Officer of the Inclusive Society Institute. This article draws on a presentation he made earlier this year at the Türkiye – Africa Media Forum in Istanbul, and suggests prioritisation of the issue at the upcoming G20 in Johannesburg later this year.