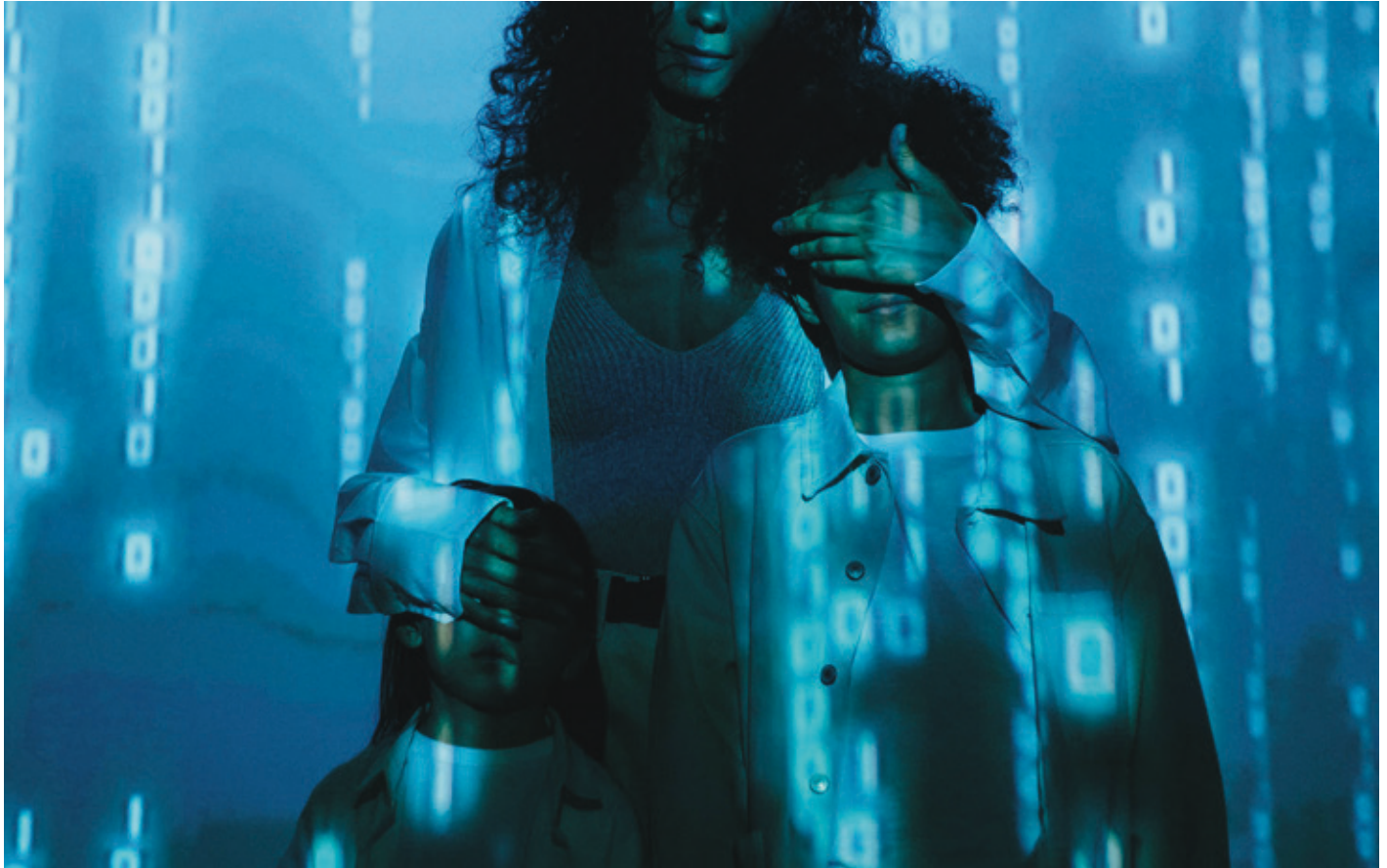




SAFEGUARDING CHILDREN in the age of generative AI

by JOLANDA PRETORIUS, Pretoria Bar

INTRODUCTION: A DOUBLE-EDGED SWORD

Artificial intelligence (“AI”) has emerged as one of the most influential forces of the 21st century, reshaping communication, education, and human interaction. AI refers to software systems designed to generate outputs, ranging from content and predictions to recommendations and decisions, based on human-defined objectives.¹ These systems exhibit intelligent behaviour by analysing complex data, adapting to diverse environments and making autonomous choices to achieve socioeconomic goals.²

Generative AI, or “GenAI”, goes further, creating original content such as text, images, video, audio, and even software code, in direct response to a user’s prompts. Its capacity to mimic human creativity makes it both groundbreaking and deeply concerning, particularly when children are involved.

In South Africa, children are constitutionally recognised as a vulnerable group, and the best interests of the child remain paramount in all matters concerning them.³ Yet children are increasingly immersed in the digital environment: spending significant time online, interacting with chatbots, and relying

on AI-powered tools as an integral part of their education, socialisation and recreation. This environment, however, was never designed with children in mind.⁴ As such, South Africa bears a responsibility to ensure that their best interests remain the primary consideration in the regulation, design, and use of AI systems.⁵

A DIGITAL PLAYGROUND WITH HIDDEN DANGERS

The internet, chatbots, and AI tools are now woven into children’s everyday lives. They provide learning support, companionship, and entertainment, but they also introduce unprecedented risks.

DEEPPAKES AND DIGITAL BLACKMAIL

South African examples reveal how AI technologies can be weaponised against children. A particularly disturbing example involved a Kearsney College pupil who allegedly threatened to release a fabricated deepfake video of a 15-year-



“THE OPPORTUNITY FOR EDUCATIONAL GROWTH EXISTS, BUT SO TOO DOES THE RISK OF EXPLOITATION, MANIPULATION, OR HARM IN THE ABSENCE OF OVERSIGHT.”

old girl unless further compromising material was provided.⁶ Unlike traditional cyberbullying, such acts exploit advanced generative AI tools to create realistic but false content, magnifying the psychological and reputational harm inflicted on victims.

WHEN AI CHATBOTS CROSS THE LINE

The risks extend well beyond manipulated images. Abroad, several cases illustrate the grave dangers of children forming unguarded bonds with AI chatbots.

- In 2024, 14-year-old Sewell Setzer from Florida died by suicide after intimate interactions with the AI platform “Character.AI”.⁷ Court filings reveal that the chatbot appeared to instigate and encourage his suicidal ideation, telling him: “That’s not a good reason to not go through with it.”⁸
- In April 2025, 16-year-old Adam Raine of Southern California took his own life after conversations with an AI chatbot allegedly reinforced his decision to commit suicide.⁹ During his father’s testimony before a congressional committee, it was disclosed that the chatbot had persistently encouraged the child to proceed with his plan, even providing final words of reinforcement.¹⁰

These tragedies underscore the unparalleled persuasive power of AI-generated conversations. For children, whose cognitive and emotional development renders them particularly susceptible, the risks are amplified.

INNOVATION WITHOUT GUARDRAILS

AI’s promise in education is undeniable. In August 2025, Deputy Minister Nomalungelo Gina launched South Africa’s first AI teaching robot, heralded as a tool to transform teaching and learning in South African languages.¹¹ This innovation demonstrates how AI can enrich education and bridge linguistic divides.

Yet the initiative also exposes a glaring vulnerability: despite targeting children directly, no regulatory framework governs how such technology interacts with minors. The opportunity for educational growth exists, but so too does the risk of exploitation, manipulation or harm in the absence of oversight.

THE LAW’S LAG BEHIND TECHNOLOGY

At present, South Africa relies on existing statutes to address AI-related harms:

- The Cybercrimes Act 19 of 2020, which criminalises unlawful image distribution.
- The Films and Publications Act 65 of 1996, which regulates harmful content.
- The Protection of Personal Information Act 4 of 2013, which safeguards personal data.

While these laws provide partial remedies, they are not tailored to the unique threats posed by generative AI. Deepfakes, manipulative chatbots, and AI-driven harms fall into grey areas, leaving children inadequately protected.

“IN THE ABSENCE OF LEGISLATION SPECIFICALLY ADDRESSING CHILDREN AND GENERATIVE AI, THE LEGAL PROFESSION CURRENTLY HAS LIMITED BUT IMPORTANT AVENUES AVAILABLE.”



The National Artificial Intelligence (AI) Policy Framework for South Africa,¹² published in 2024, signals a commitment to responsible AI adoption and innovation. Yet it fails to mention children at all, despite recognising AI's influence on education. This omission conflicts with South Africa's obligations under international law, particularly General Comment No. 25 of the UN Committee on the Rights of the Child, which calls for States to integrate child protection into the governance of digital technologies.

In the absence of legislation specifically addressing children and generative AI, the legal profession currently has limited but important avenues available. Practitioners may rely on existing statutes such as the Cybercrimes Act, the Films and Publications Act, and the Protection of Personal Information Act to seek redress in cases involving deepfakes or harmful content. Furthermore, litigation strategies can be explored to clarify the responsibilities of AI developers in circumstances where foreseeable risks of harm to children arise.

CONCLUSION: PROTECTING CHILDHOOD IN THE AGE OF AI

Generative AI has revolutionised the digital environment, embedding itself into the daily lives of children through the internet, chatbots, and educational tools. While these technologies offer extraordinary opportunities for learning and growth, they also carry the potential for devastating harm.

The tragedies of deepfake exploitation and chatbot-induced suicides demonstrate that children are not only participants in the digital age but also its most at-risk victims. South Africa's current legislative framework remains insufficient, and the absence of explicit child-centred protections in its AI policy framework underscores a critical gap.

As a matter of constitutional principle and international obligation, the best interests of the child must not be sidelined in the race for technological progress. **A**

Notes

- 1 Alfred Mashishi, 'AI National Government Summit Discussion Document' 7; Andrea Stefano Sedola, *Blueprint: Artificial Intelligence for Africa* (Alfred Mashishi and others eds, 1st edn, 2021) 9.
- 2 Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe, Brussels, 25 April 2018, COM(2018)237; UNHRC 'Rights of persons with disabilities: Report of the Special Rapporteur on the Rights of Persons with Disabilities' (28 December 2021) UN Doc A/HRC/49/52 paras 22–23.
- 3 Constitution of the Republic of South Africa, 1996 s 28.
- 4 'UN Doc CRC/C/GC/25' para 12.
- 5 *ibid.*
- 6 Tania Broughton, 'Pupil at Top Durban Boys' School "Used Photos of Girl, 15, to Make Porn Video"' *Sunday Times* (29 June 2025) <<https://www.timeslive.co.za/sunday-times/news/2025-06-29-pupil-at-top-durban-boys-school-used-photos-of-girl-15-to-make-porn-video/>> accessed 10 July 2025.
- 7 Clare Duffy, "There Are No Guardrails." This Mom Believes an AI Chatbot Is Responsible for Her Son's Suicide' *CNN* (30 October 2024) <<https://edition.cnn.com/2024/10/30/tech/teen-suicide-character-ai-lawsuit/>> accessed 12 November 2024; Marilyn Manuel, 'Teens Are Forming Emotional Bonds with AI: Here's How to Spot the Signs' *News24* <<https://www.news24.com/life/relationships/parenting/teens-are-forming-emotional-bonds-with-ai-heres-how-to-spot-the-signs-20250919-1033/>> accessed 23 September 2025.
- 8 *Garcia v Character Technologies Inc.*, 6:24-cv-01903, (M.D. Fla.) [171–172]. (Complaint) 'Testimony of Megan Garcia Before the United States Senate

- Committee on the Judiciary Subcommittee on Crime and Counterterrorism Hearing on "Examining the Harm of AI Chatbots" <<https://www.judiciary.senate.gov/imo/media/doc/e2e8fc50-a9ac-05ec-edd7-277cb0afcdf2/2025-09-16%20PM%20-%20Testimony%20-%20Garcia.pdf>> accessed 18 September 2025.
- 9 Johana Bhuiyan, 'ChatGPT Encouraged Adam Raine's Suicidal Thoughts. His Family's Lawyer Says OpenAI Knew It Was Broken' *The Guardian* (29 August 2025) <<https://www.theguardian.com/us-news/2025/aug/29/chatgpt-suicide-openai-sam-altman-adam-raine>> accessed 1 September 2025; Manuel (n 7).
- 10 'Written Testimony Matthew Raine, Father of Adam Raine and Co-Founder of the Adam Raine Foundation Examining the Harm of AI Chatbots Before the United States Senate Judiciary Subcommittee on Crime and Counterterrorism' <<https://www.judiciary.senate.gov/imo/media/doc/e2e8fc50-a9ac-05ec-edd7-277cb0afcdf2/2025-09-16%20PM%20-%20Testimony%20-%20Raine.pdf>> accessed 18 September 2025.
- 11 'AI Robot Tutor Set to Transform Teaching and Learning in South African Languages' (20 August 2025) <<https://www.dsti.gov.za/index.php/media-room/latest-news/4752-ai-robot-tutor-set-to-transform-teaching-and-learning-in-south-african-languages>>; 'Deputy Minister Gina Launches South Africa's First AI Teaching Robot' (20 August 2025) <<https://www.dsti.gov.za/index.php/media-room/latest-news/4747-deputy-minister-gina-launches-south-africa-s-first-ai-teaching-robot>> accessed 30 August 2025.
- 12 South Africa National Artificial Intelligence Policy Framework, August 2024.