Check for updates

Hauntings and hallucinations

As I write this Leader, I am currently dealing with a dispute over material submitted to our Journal which we believe contains hallucinated references and other AI-generated material. As if by magic, as I was writing the first sentence, an email popped up on my screen with a heading which begins "Leslie, ready to navigate the future of AI and unlock new opportunities…?", and offering me (for a fee) the chance to participate in a course which will help me, and I quote, "to build the skills needed to confidently navigate the rapidly evolving world of Artificial Intelligence and understand its impact on business, leadership, and the future of work". The course, in fact, looks good, and comes from a reputable source, but I am interested here in what I am learning from my own reaction to this solicitation coming hot on the heels of our latest dispute over AI-generated material. My reaction, I am embarrassed to say, is along the lines of "Go away and leave me alone".

At the *South African Journal of Science*, we have tried to be, and continue to try to be, agile about AI and large language models (LLMs). We have a policy which, although always subject to change (and we are in the process of tweaking it), we believe is helpful, and our most recent Guest Leader summarises where we are and argues helpfully, we believe, for how to ensure responsible use. We are also lucky to be the Journal of the Academy of Science of South Africa (ASSAf), which has coauthored useful guidelines and hosts regular engagements on the issues at stake. Why, then, do I (and many others) feel so depleted about the (often fascinating) issues surrounding AI use?

A short answer to this is that journals like ours are facing unprecedented pressures, of which this is one example. For example, our own rate of submissions has skyrocketed, largely through the dramatic increase in submissions from Asia which do not fit our criteria for interdisciplinary work from and/or about Africa. It is both time-consuming and depressing to have to consider, in the interests of fairness, submissions of varying quality (from excellent to very poor) which simply, for reasons of scope, cannot be considered for our Journal. We are looking to develop AI methods to at least sort submissions into 'probably out of scope' versus 'probably in our scope', and this should help us, but we are not there yet, and even when we are there, according to our own policies, AI sorting suggestions must still be overseen at some point by humans. Unlike Hogwarts School in the Harry Potter books, we do not have a magic sorting hat. As a person who is interested in AI but not an AI expert by any stretch of the imagination, I listen with interest to corridor talk about articles having been written by bots, and about bots submitting to journals articles written by humans. I do not have the time or skills to check whether there is a good evidence base for these and other techno-conspiracy theories, but I am fascinated by them. Depending on how you define science, and especially if, like me, you value rigorous, evidence-based science, this is hardly very scientific of me, but it is, if I may suggest this, human.

Some of the context of this, I think, is one in which things are happening which are fundamentally disturbing, and not for the most obvious reasons. The term 'catfish' has only recently taken on a meaning which has nothing to do with actual fish. According to the *Oxford English Dictionary*, since 2023, a secondary official meaning of the word 'catfish' is "a person who deceives by means of a fictional or assumed online persona, esp. with the intent to lure someone into a relationship". The usage comes from the 2010 film *Catfish*[1] and the subsequent spin-off television series. Victims of catfishing fall in love with made-up personas, sometimes even sending money and gifts to online fake lovers. Catfishing as a practice has been enabled by technology, but the abuse catfishes perpetrate is deeply and timelessly human – we cheat, we lie, we extort. Part of the many evidence-based traditions in science, in fact, were put in place long before the Internet was invented precisely to minimise the possibility of subjectivity in scholarship, with considerable attention paid to ways in which scientists can interrogate and deal with, as far as is possible, our own unconscious biases, of which we may not be aware.

Recently at our Journal, we received a submission written by a very senior and prominent academic and this submission used hallucinated references (references which do not exist but are AI-generated, also known as hallucitations). We also received a 'peer review report' which is clearly just the result of a peer reviewer we ourselves appointed (also a senior academic) uploading a submission to AI and asking AI to generate comments. I confess to having felt enraged by these occurrences. There were, of course, genuine concerns at stake. For example, we worry as a journal that, despite our best efforts, we enabled by proxy the work of an author to have been uploaded onto an AI platform without the author's permission. But I think my own rage at these sorts of infractions may well be out of proportion to the legitimate or surface-level concerns. For me (and possibly for many of us), part of the anger has to do with concerns not about technology itself but about how it enables a world in which dearly held, taken-for-granted rules of science and ethics – the bedrock of a shared and valued identity – can be violated not by 'the next generation of scholars' (not that this is not a concern) but by colleagues whom we trusted by virtue of their positions. According to the developmental psychologist Erik Erikson[2], the establishment of trust versus mistrust in others is the most fundamental stage in human development (0–18 months); when trust which has been taken for granted for a long time is violated, the threat may be quite fundamental to how we see ourselves professionally, but also personally. And of course, if others can break rules in what appear to be quite flagrant ways, those of us observing and reacting may also become uncomfortably reacquainted with the parts of ourselves which would also like to break rules and get away with it.

In my (happy and very privileged) years as Editor-in-Chief of this interdisciplinary scientific journal, I have deliberately not emphasised or leaned on my training as a clinical psychologist to make the kinds of speculative comments I have indulged myself in here. As far as I am aware, though, I have good motives for breaking my rule, to some extent, now. I do not think that the scholarly community in our context has sufficiently considered the psychological impact of the violation of rules enabled by AI and LLMs, and the corrosive worry that the only perpetrators we catch may well be those who are the less skilled perpetrators. The issues of psychological responses to abuse of AI are, like research into mental health responses to climate change, empirically researchable. And in neither case is the only issue the fear of machines taking over, or unimagined new environments which we fear. Another issue has to do with fundamental questions we are implicitly asking ourselves about what it means to be human. These issues are not new, and it is not surprising that some contemporary psychoanalytically informed research, including research focusing on colonial and post-colonial issues[3], explores historical concerns from within the lens of what is commonly termed 'hauntology'. One issue which has not been considered sufficiently is the intertwining of AI use with the reproduction, especially relevant in our context, of the haunting legacies of colonialism in our work. At this time, when AI writes and hallucinates for us, making our work smoothly and superficially fluent, the language it uses is not generally the language or languages of the Global South. There is so much to consider and learn from in the responsible use of AI, but at least some of the 'ghost in the machine' that we may fear with AI, is probably not in the machine but in us. For very human reasons, we need to work together across disciplines and in recognition of our anxieties, at this time more than ever.

References

1. Joost H, Schulman A. Catfish [film]. Los Angeles, CA: Rogue Pictures; 2010.
2. Erikson E. Childhood and society. New York: W. W. Norton; 1950.
3. Deumert A. Thinking, living and acting with ghosts: Hauntology and Fanon's zone of occult instability. Emotion Space Soc. 2026;59, Art. #101172. https://doi.org/10.1016/j.emospa.2026.101172

HOW TO CITE:
Swartz L. Hauntings and hallucinations. S Afr J Sci. 2026;122(7/8), Art. #27632. https://doi.org/10.17159/sajs.2026/27632