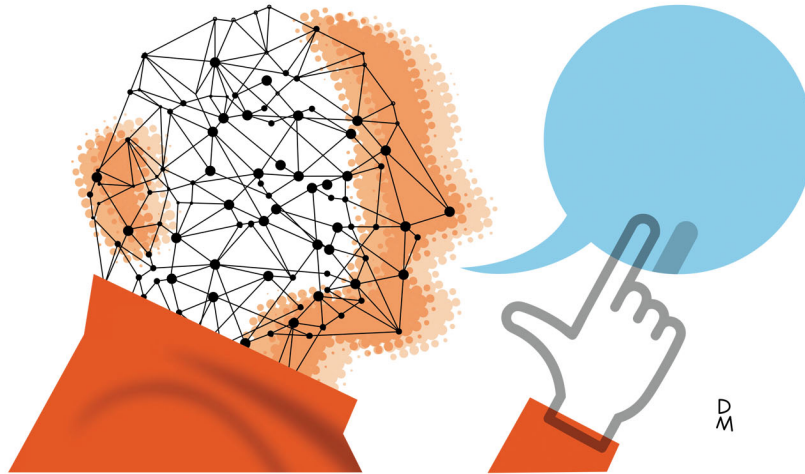


Ground rules for ethical AI use needed to be successful innovation hub



Here are three suggested criteria for the use of large language models.

Sebastian Porsdam Mann, Brian D. Earp and Julian Savulescu

After a severe bout of Covid-19, a colleague, Sumeeta, found herself facing an unexpected challenge. Despite her intact verbal intelligence and reasoning skills, she suddenly struggled with the mechanics of writing. Constructing grammatical sentences and coherent paragraphs now became an exhausting ordeal. A single paragraph required a full hour of recovery time. This experience threatened to derail her academic career.

Then she discovered the power of large language models (LLMs). For Sumeeta, LLMs like ChatGPT or Claude became a lifeline, enabling her to once again engage in writing tasks that had seemed impossible post-illness. From personal correspondence to academic essays, these AI tools helped bridge the gap between her ideas and their expression.

Her story is a reminder of the potential of AI for good when used responsibly and creatively. And indeed, Singapore has made significant strides in encouraging responsible use of LLMs in universities. A recent study at the National University of Singapore (NUS) found that 76 per cent of students use AI tools like ChatGPT to summarise texts, while 71.8 per cent employ them to collect information and formulate ideas.

In a nod to the potential of these powerful tools, NUS recently issued guidance for students on the ethical use of AI tools, emphasising transparency, academic integrity, acknowledgement and personal responsibility for submitted work.

But students are not the only important group using these language models. Another recent survey estimated that 40 per cent of Singaporean workers are already using AI in their jobs, with over 90 per cent reporting increased productivity. Yet the study also found that 76 per cent of these AI users admit to passing off AI-generated work as their own. And while scandals involving misuse of these models – for example, lawyers submitting documents based on made-up cases – seem to be rare in Singapore so far, these stories from abroad are worrying.

How can we harness the good that LLMs can bring, as in the example of Sumeeta, while avoiding misuse? We need to make sure we include another key segment: not just students, but also researchers, academics and other professional knowledge workers across our economy.

'OUTCOME GOODS' AND 'PROCESS GOODS'

The stakes are high. Singapore's position as a global hub for innovation and research can benefit immensely from AI use, not just in terms of quantity but also quality. Yet maintaining Singapore's lauded status depends on ensuring the highest standards of intellectual integrity. Sometimes a distinction is made between "outcome goods" and "process goods" in the context of knowledge work and research. Outcome goods are end results we're aiming for: generating reliable, innovative knowledge and understanding, equipping us to solve real-world problems.

Process goods are about the integrity of how we achieve those outcomes: here the focus is on means, not only ends. These goods include things like fairness, transparency, and adherence to ethical and legal standards. LLMs have the potential to significantly enhance Singapore's outcome goods. They can accelerate research, streamline knowledge work, and even democratise access to information and idea generation.

For all these reasons, their responsible use should be encouraged in workplaces, universities and government offices. Sumeeta's experience is a testament to how these tools can unlock real potential and enable valuable contributions that might otherwise be lost.

However, the potential for widespread misuse of LLMs in the workplace suggests that we're at risk of compromising our process goods. When individuals copy and paste generated text without checking it, or pass off AI-generated work as their own, this has serious impacts on trust and fair competition, and may devalue the human creativity and critical thinking that are essential to successful innovation.

In a recent research paper with colleagues from Harvard, Cambridge and Copenhagen, published in the latest issue of *Nature Machine Intelligence* we argue that policies for LLM use need to address both process goods and outcome goods. We shouldn't just aim to maximise useful output; it matters how we do this as well. Appropriately

balancing both types of goods will be necessary to maximise the benefits and minimise the risks and harms – including harms to personal integrity, and to Singapore's reputation more broadly – in adopting these powerful technologies.

GUIDELINES FOR ETHICAL USE

In our article entitled "Guidelines for ethical use and acknowledgment of large language models in academic writing", we propose three criteria for the ethical use of LLMs in professional and academic settings: human vetting/guaranteeing, substantial human contribution, and acknowledgment/transparency.

Let's explore what these mean and why they matter for Singapore's knowledge workers. Since LLMs are known to make mistakes or invent information, human vetting and guaranteeing should be non-negotiable, especially in fields where accuracy and reliability are paramount.

Imagine a policy analyst at a government ministry using an LLM to help them analyse global economic data and generate initial policy recommendations. While the AI might efficiently compile and synthesise information, the analyst commands the expertise in Singapore's unique economic position, understanding of local business ecosystems, and ability to anticipate geopolitical implications. His role would be crucial for verifying the accuracy and relevance of any LLM suggestions to the Singapore context, which will be necessary for crafting reliable and effective policies.

The second criterion, substantial human contribution, serves two crucial purposes: ensuring human expertise enhances AI-generated work and providing a basis for fair credit attribution.

Consider a data scientist at a Singapore-based fintech startup using an LLM for market analysis. While the AI processes vast data, the scientist must critically evaluate its suggestions, incorporating local market knowledge and factors like regulatory changes. Simply passing off an AI analysis as one's own is problematic: it's unfair to colleagues doing original work; misrepresents individual skills, potentially leading to undeserved rewards; and risks an overreliance on AI that could erode essential human analytical skills.

The third criterion, acknowledgment and transparency, is crucial for maintaining public trust and

ensuring accountability. Consider an enforcement authority (or agency) using an LLM to assist in processing Employment Pass applications. Transparency about this AI use is vital. If an application is rejected, the applicant needs to know AI was involved to understand how their data was processed and potentially appeal over the decision.

Education is another key example. Students should be allowed to use LLMs, but must be required to be transparent about the nature and extent of its use. This allows teachers to accurately assess students' abilities and tailor their instruction accordingly.

It's important to note that these three criteria – human vetting, substantial contribution and acknowledgment – will manifest differently across various sectors and projects. Yet they touch on areas that are likely to be important, to various degrees, across all knowledge work.

For academics, we've proposed a specific, standardised acknowledgement statement for research publications: "Any use of generative AI in this manuscript adheres to ethical guidelines for use and acknowledgment of generative AI in academic research. Each author has made a substantial contribution to the work, which has been thoroughly vetted for accuracy, and assumes responsibility for the integrity of their contributions."

We argue that additional details on LLM use and impact should also be added where necessary to allow reproducibility of research and judgments of research credibility by independent experts.

As Singapore continues to invest heavily in AI, the goal should be to maximise long-term benefits through responsible use. This requires open, informed discussions across all sectors to develop appropriate guidelines that balance innovation with integrity. It also requires investment in new skills.

Employers, educational institutions and the government should consider providing learning opportunities focused on both the ethical considerations and practical applications of AI tools. Just as Singapore has invested in digital literacy in the past, the time has come to invest in AI literacy for all knowledge workers.

Sebastian Porsdam Mann is visiting research fellow, Brian D. Earp is associate professor and Julian Savulescu is Chen Su Lan Centennial Professor of Medical Ethics at the National University of Singapore.