

With the far-reaching impact that artificial intelligence has on our everyday lives, companies' AI ethics teams bear the colossal burden of ensuring that this technology is safe and fair, say the writers.

PHOTO: ISTOCKPHOTO



# Moving AI ethics beyond guidelines

The recent row over a leading researcher's departure from Google raises questions about the role and remit of AI ethics teams and the inadequacy of frameworks to deal with moral complexities

**Lim Sun Sun and Jeffrey Chan Kok Hui**

For The Straits Times

The departure of artificial intelligence (AI) researcher Timnit Gebru from Google under controversial circumstances has raised discomfiting questions about the company's stance on AI ethics. It has also revealed the challenges of practising AI ethics on the front line of this field.

Dr Gebru, who was co-leader of Google's Ethical AI team, had earned widespread acclaim for her earlier work highlighting that AI facial recognition was less capable of identifying women and people of colour, thereby perpetuating discrimination if unchecked.

Her alleged dismissal from Google was apparently triggered by her latest research paper questioning inherent biases in large models used to train algorithms for

language processing. She also highlighted the staggering environmental costs of training such models, given the considerable computer processing power and electricity involved.

She cited previous research that had found that training one language model generated as much carbon dioxide as the lifetime output of five average American cars.

Dr Gebru asserts that she was pressured by higher-ups in the company to retract the paper from a forthcoming research conference or to remove the Google employees' names from it.

In response, chief executive Sundar Pichai stated in a company-wide memo that Google should seek to improve the processes leading to her dismissal and framed it as a failure to protect the rights of a black, female minority employee, but did not address the issue of her research being censored.

Thousands of Google employees and individuals from other

organisations have since endorsed an open letter expressing support for Dr Gebru and the public pressure continues to mount.

## DEALING WITH DILEMMAS

This unfortunate episode that is far from resolved holds interesting lessons for AI ethics. That a technology behemoth such as Google even has an AI ethics team is noteworthy in and of itself.

It underscores how society's intensifying deployment of AI has unleashed an expanding litany of ethical dilemmas around automation, datafication and surveillance that technology companies must grapple with.

While it is taken for granted that large companies must have finance, legal, marketing and human resource departments, our technologising world does indeed necessitate that companies also hire ethics teams to provide guidance on issues relating to moral responsibility and civic duty.

But this then begs the question as to the roles and remit of such ethics teams.

Given that ethics is about the morally good life, one that is to be reflected in our AI milieu, then the crucial matter of how to define the organisational role, discretions and safety net of professional ethicists taking after Dr Gebru remains an outstanding task.

With the far-reaching impact AI

has on our everyday lives, AI ethics teams bear the colossal burden of ensuring that this technology is safe and fair.

AI-powered algorithms increasingly make many high-stakes decisions with potentially serious consequences for lives and society – from meting out legal penalties to qualifying for a loan to landing a job.

While AI technologies present clear benefits, they can nevertheless bring about different harms. These harms do not only include the direct harms manifested in malicious adversarial attacks and disinformation, they also extend to the indirect harms perceived when organisations and societies fail to check data biases or nuanced discrimination when using AI tools.

AI ethics teams like Dr Gebru's must therefore weigh the benefits and harms introduced by AI processes so as to flag immediate implications for their company, but also to caution against long-term repercussions for humanity at large.

In practice, therefore, if such ethics teams are to be more than a token of the company's corporate social responsibility, are they to serve as the proverbial conscience of the organisation and rein it in when it wades into ethical grey areas?

Or is their job to educate colleagues on the potential ethical pitfalls they could land in, and thereby imbue in their engineers and designers an instinctive appreciation for their ethical burdens? Or perhaps their key function is to develop ethical guidelines for the organisation as it forges groundbreaking innovations without ethical precedents, and then to clarify and settle ethical conflicts that may result?

## INADEQUATE MODELS

Given that there is in fact no lack of AI ethics guidelines or model frameworks

today. In an important study evaluating AI ethics guidelines, Dr Thilo Hagendorff from the University of Tuebingen in Germany counted at least 22 major ethical guidelines in the world.

And this number is surely set to rise with the recent introduction of the Cyberspace Administration of China's guidelines on data collection and Singapore's evolving AI Ethics And Governance Body Of Knowledge framework.

However, criticisms of such ethics guidelines also abound.

They range from their ineffectiveness because of inadequate enforcement, to the neglect of feminist ethical principles of care and ecological concerns when developing these guidelines. Furthermore, stating clear ethical principles and values upfront does not always result in unambiguously ethical outcomes.

Consider an example from the European Commission's influential Ethics Guidelines For Trustworthy AI published last year. Four ethical principles, namely, "Respect for human autonomy", "Prevention of harm", "Fairness" and "Explainability", undergird this guideline.

Nevertheless, to prevent harm sometimes, human autonomy may have to be violated – for instance, when predictive policing aims to reduce crime through constant surveillance that impinges on individual privacy and freedom.

These guidelines neither inform AI developers of how to translate ethical principles into mathematical functions, nor how to make the most ethical trade-off between contesting principles in their models.

In other words, these guidelines cannot settle conflicts of ethical principles and values when they clash. Only individuals and organisations willing to embody these ethical guidelines, and to transform them into actionable thoughts and deeds, can do that.

These are tasks that AI ethics teams alone cannot undertake, especially if they are not accorded some modicum of protection and security when drawing out inconvenient truths and imposing constraints that no conscientious organisation should violate.

## BUILDING ETHICAL SCAFFOLDS UPSTREAM

Building robust ethical scaffolds upstream is another urgent endeavour to pursue.

Principally, we must ensure that our next generation of technology professionals are fully cognisant of the moral complexities of their work.

They must learn to appreciate how their apps, codes, programs, software and structures can have large social impacts beyond their technological applications.

They must also learn how to integrate and amplify principles of beneficence, fairness, justice and transparency in their designs.

At the Singapore University of Technology and Design, we train our students to navigate the rich but also chequered terrains of ethics. At the end of their first year, undergraduates are required to take a mandatory course on ethics as part of the Professional Practice Programme.

This course serves as a primer for more advanced humanities, arts and social sciences electives on AI ethics from such diverse disciplines as anthropology, design theory, history and philosophy.

The aim is to continuously and progressively buttress students' familiarity with and understanding of ethics, so that they can be ready to take on the complex moral challenges presented by AI practices in their professional lives.

## CORPORATE ACCOUNTABILITY

We must also complement such educational interventions by moving decisively from AI ethics guidelines to considering regulations that hold technology companies accountable to concrete ethical standards.

For example, under Singapore's Resource Sustainability Act that introduced the Extended Producer Responsibility approach, electrical and electronic goods manufacturers are now legally obligated to collect and treat the e-waste their products generate when they reach end-of-life.

Similarly, technology companies should also be subject to regulations governing carbon footprint thresholds for computing processes that power AI-driven solutions.

The salutary discourse around the promise of AI must be grounded in a recognition of the possible harms it can wreak.

While ethics teams and guidelines are steps in the right direction, they risk being trampled upon in the race for technological domination.

stopinion@sph.com.sg

- Lim Sun Sun is professor of communication and technology and head of humanities, arts and social sciences. Jeffrey Chan Kok Hui is assistant professor of design theory and ethics. They are both faculty members at the Singapore University of Technology and Design.

# Streaming shrinks the Hollywood star

**John Gapper**

Hollywood thrives on loud, colourful confrontations in films such as last year's *Avengers: Endgame* and the forthcoming *Godzilla Vs Kong*. So directors and actors have not held back since WarnerMedia declared that it would suspend nearly a century of tradition in the face of pandemic disruption.

Warner is to release all of its 17 new films next year on its streaming service HBO Max at the same time as in United States cinemas.

Cue outrage, with director Christopher Nolan firing back last week that the stars "went to bed the night before thinking they were working for the greatest movie studio and woke up to find out they were working for the worst streaming service".

Ouch. Nolan has a way with words, as well as with fitting action around baffling conceptual plots in films such as *Inception* and this year's *Tenet*.

He might well be annoyed: Since AT&T took over Time Warner

(including the Warner Bros studio and HBO) in 2018, it has focused intently on boosting broadband and mobile phone subscriptions, rather than stroking the talent.

More than cinematic pride lies behind the anger at films such as the science fiction epic *Dune* becoming subscription fodder.

Stars, agents and directors depend for their wealth on sharing in box-office profits, as well as being paid a fee. "Ultimately, people want to make money,"

Mr John Stankey, AT&T chief executive, observed last week.

But Nolan is surely correct that his streaming overlords "don't even understand what they're losing" in the effort to catch Netflix and Disney Plus, their more successful rivals.

Hollywood's most profitable films are cultural and financial blockbusters that are shrunk by the small screen.

It does not mean that television is inferior. Last week, I watched *Man, David Fincher's film about Herman Mankiewicz, who wrote the screenplay of Citizen Kane with Orson Welles. Some of the ravishing black and white cinematography*



Hugh Grant and Nicole Kidman, stars of the HBO show *The Undoing*, would not have been seen dead on the small screen a few years ago, but the growth of streaming has made performers less snooty about television, says the writer.

PHOTO: HBO

to seize the opportunity to shake up Hollywood.

Disney is releasing Pixar's *Soul* to Disney Plus's 74 million subscribers on Christmas Day in the US, so Warner is not alone.

But it looks a bit desperate. HBO Max was a flawed idea – taking a pioneering service with a first-rate reputation in HBO and bundling it confusingly with other Warner content. Having gained only 13 million active subscribers, it now wants to pour new films down the same funnel. No wonder the talent is irked.

Films such as *Dune* and *Godzilla Vs Kong*, with US\$150 million-plus budgets, are not made for television. They are produced to be events, with huge marketing campaigns and a series of release windows from cinema to streaming to keep the excitement, and the revenue streams, bubbling away for years. It may not be art, but it is a craft.

Disney is expert at managing its franchises, turning cartoon princesses into dolls and adventure films into theme park rides. Its space western hit, *The Mandalorian*, on Disney Plus is a spin-off from the Star Wars films.

Let cinema atrophy and the entire entertainment edifice would suffer.

Warner intends to return to US cinema-first releases when the pandemic passes. But the writing is on the wall: If streaming takes over entirely, the blockbuster as we know it becomes an endangered species. No doubt someone will make a movie about it one day. FINANCIAL TIMES

would have shone in cinemas where it was released first, but its intimacy suited television.

Much drama of the kind that used to be a mainstay of cinema has migrated to television, thanks to HBO and Netflix. That is where today's successors to Mankiewicz have become auteur showrunners of series such as *The Sopranos* and *The Crown* – while film directors make action flicks, romantic comedies and young adult epics.

Since the *Star Wars* and *Superman* films of the 1970s, Hollywood has gone as big as possible to lure people from their homes to pay for the communal experience of watching

spectacular action, complete with booming sound, at cinemas. It has aimed at the widest audience, not the connoisseur.

Globalisation has helped the formula to work: Global box office revenues rose to US\$42 billion (\$56 billion) last year. Teenagers are faithful to long-running film franchises such as those featuring Disney's Marvel superheroes; 12- to 17-year-olds in the US and Canada went to the movies an average of 4.9 times last year, compared with 2.9 for people in their 50s.

The cinema is under most strain in the US, where box-office revenues fell last year, even before the pandemic (the US and Canada

comprise only about one-quarter of global box-office receipts).

Nolan's *Tenet* was finally released in Britain in August in an effort to get the audience back, but it struggled in US cinemas.

The growth of streaming has already made performers less snooty about television. Nicole Kidman and Hugh Grant, stars of the HBO show *The Undoing*, are A-list actors of the kind who would not have been seen dead on the small screen only a few years ago.

It is easy to see the temptation for Warner, facing a pile-up of new releases amid the pandemic and a streaming service in need of help,