

Opinion

SUNDAY, NOVEMBER 3, 2019

MELANIE MITCHELL

INTELLIGENT MACHINES CATASTROPHICALLY misinterpreting human desires is a frequent trope in science fiction, perhaps used most memorably in Isaac Asimov's stories of robots that misconstrue the famous 'three laws of robotics.' The idea of artificial intelligence going awry resonates with human fears about technology. But current discussions of superhuman AI are plagued by flawed intuitions about the nature of intelligence.

We don't need to go back all the way to Isaac Asimov — there are plenty of recent examples of this kind of fear. Take a recent op-ed essay in *The New York Times* and a new book, *Human Compatible*, by the computer scientist Stuart Russell. Dr Russell believes that if we're not careful in how we design artificial intelligence, we risk creating 'superintelligent' machines whose objectives are not adequately aligned with our own.

As one example of a misaligned objective, Dr Russell asks, "What if a superintelligent climate control system, given the job of restoring carbon dioxide concentrations to preindustrial levels, believes the solution is to reduce the human population to zero?" He claims that "if we insert the wrong objective into the machine and it is more intelligent than us, we lose."

Dr Russell's view expands on arguments of the philosopher Nick Bostrom, who defined AI superintelligence as 'an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills.' Dr Bostrom and Dr Russell envision a superintelligence with vast general abilities unlike today's best machines, which remain far below the level of humans in all but relatively narrow domains (such as playing chess or Go).

Dr Bostrom, Dr Russell and other writers argue that even if there is just a small probability that such superintelligent machines will emerge in the foreseeable future, it would be an event of such magnitude and potential danger that we should start preparing for it now. In Dr Bostrom's view, 'a plausible default outcome of the creation of machine superintelligence is existential catastrophe.' That is, humans would be toast.

These thinkers — let's call them the 'superintelligentsia' — speculate that if machines were to attain general human intelligence, the machines would quickly become superintelligent. They speculate that a computer with general intelligence would be able to speedily read all existing books and documents, absorbing the totality of human knowledge. Likewise, the machine would be able to use its logical abilities to make discoveries that increase its cognitive power.

Such a machine, the speculation goes, would not be bounded by bothersome human limitations, such as slowness of thought, emotions, irrational biases and need for sleep. Instead, the machine would possess something like a 'pure' intelligence without any of the cognitive shortcomings that limit humans.

The assumption seems to be that this AI could surpass the generality and flexibility of human intelligence while seamlessly retaining the speed, precision and programmability of a computer. This imagined machine would be far smarter than any human, far better at 'general wisdom and social skills,' but at the same time it would preserve unfettered access to all of

Should we fear superintelligence?

AI resonates with human fears about technology but superintelligence is a flawed concept and shouldn't inform our policy decisions



A robot welcomes participants to the Future Investment Initiative forum in Riyadh, Saudi Arabia

its mechanical capabilities. And as Dr Russell's example shows, it would lack humanlike common sense.

The problem with such forecasts is that they underestimate the complexity of general, human-level intelligence. Human intelligence is a strongly integrated system, one whose many attributes — including emotions, desires, and a strong sense of selfhood and autonomy — can't easily be separated.

Similarly, if generally intelligent AI is ever created (something that will take many decades, if not centuries), its objectives, like ours, will not be easily 'inserted' or 'aligned.' They will rather develop along with the other qualities that form its intelligence, as a result of being embedded in human society and culture. The machines' push to achieve these objectives will be tempered by the common sense, values and social judgment without which general intelligence cannot exist.

What's more, the notion of superintelligence without humanlike limitations may be a myth. It seems likely to me that many of the supposed deficiencies of human cognition are inseparable aspects of our general intelligence, which evolved in large part to allow us to function as a social group. It's possible that the emo-

tions, 'irrational' biases and other qualities sometimes considered cognitive shortcomings are what enable us to be generally intelligent social beings rather than narrow savants. I can't prove it, but I believe that general intelligence can't be isolated from all these apparent shortcomings, either in humans or in machines that operate in our human world.

In his 1979 Pulitzer Prize-winning book, *Gödel, Escher, Bach: an Eternal Golden Braid*, the cognitive scientist Douglas Hofstadter beautifully captures the counter-intuitive complexity of intelligence by posing a deceptively simple question: "Will a thinking computer be able to add fast?" Dr Hofstadter's surprising but insightful answer was, 'perhaps not.'

As Dr Hofstadter explains: "We ourselves are composed of hardware which does fancy calculations but that doesn't mean that our symbol level, where 'we' are, knows how to carry out the same fancy calculations. Let me put it this way: There's no way that you can load numbers into your own neurons to add up your grocery bill. Luckily, your symbol level (i.e., you) can't gain access to the neurons which are doing your thinking — otherwise you'd get addlebrained." So, why, he asks, "should it not be the same

for an intelligent program?"

In other words, the intelligent part of your mind can't harness the fast-adding skills of your own neurons, and for good reason. This barrier — between the "self" that you are aware of and the detailed activity of your brain — permits the kind of thinking that matters for survival without getting overwhelmed ("addlebrained") by your own thought processes. Similarly, a thinking computer's hardware, like ours, would presumably include circuits for fast arithmetic, but at the level of its cognitive awareness, the machine wouldn't be able to tap into these circuits any more than we humans can. It's fine to speculate about aligning an imagined superintelligent — yet strangely mechanical — AI with human objectives. But without more insight into the complex nature of intelligence, such speculations will remain in the realm of science fiction and cannot serve as a basis for AI policy in the real world.

Understanding our own thinking is a hard problem for our plain old intelligent minds. But I'm hopeful that we, and our future thinking computers, will eventually achieve such understanding in spite of — or perhaps thanks to — our shared lack of superintelligence.

—NYT



OUT OF MY MIND

MEGHNAD DESAI

Govt's dozen own goals on Kashmir

ALL IS FAIR in love and war. Governments are entitled to do whatever is within constitutional limits according to the political belief of their ruling party and normal rules of conduct. Others may not like it, but then they did not win the confidence of the people.

What is unforgivable is incompetence. The saga of Article 370 began at the top as a brilliant manoeuvre. The decision to abrogate Article 370 on August 5–6 was a stunning success. Somehow the government seemed to have found a window of opportunity when, with the J&K government out of office and hence the responsibility for any change in Article 370 falling on the Central government, Prime Minister Narendra Modi and Home Minister Amit Shah found a sequence of adjustments to various articles in the Constitution to make the most profound change in Article 370.

Modi and Shah had cut the Gordian knot. The de facto situation of J&K not being truly autonomous was now legally secured. As used to happen in the past, mobile telephones were shut down, public meetings banned, curfews imposed. Now, however, we were promised that these bans would not last. The promise was that business of the government would improve and healthy and prosperous Kashmiris would be empowered. New governments would be established.

This was however not a priority. The eyes of the government were firmly fixed on Pakistan and the United Nations. The avoidance of any official censure by the UN Security Council was adroitly managed by the government.

Ninety days have passed since then. There is even now no normalcy. What is obvious is that

whatever scheme the top echelon of the government had in mind has not been delivered by the lower rungs, from the governor down.

The episode involving the European Parliamentarians shows that the government does not believe in its own propaganda that all is well. Whoever dreamt up this clumsy and transparently flawed programme should be sacked. It is not just one but a dozen own goals. It has exposed a serious gap in international diplomacy.

It has been obvious for some time that there is a serious dearth of talent in political personnel below the top two. The PMO also has not been up to the challenges that the PM sets for it, as was obvious in the demonetisation case. The Article 370 saga has been allowed to harm India's reputation thanks to a lack of foresight.

An urgent policy intervention is needed. Chanakya's sequence of *saam, daam, dand* and *bhed* seems to have been reversed. *Dand* has been used at the outset. But *bhed* has failed as the MEP (Members of European Parliament) fiasco shows. *Daam* has been promised but not delivered.

What remains is *saam*. There is a need to display confidence in the rightness of the policy if indeed it is believed and, I am sure it is, that the policy was correct and in the interest of J&K.

Relax the curfew completely, release all prisoners, face the crowds of protesters and show the world that they remain a minority. Allow anyone and everyone to visit Kashmir as indeed now, after the abrogation of Article 370, they have the right. Let the world come openly rather than report furtively. Somehow people believe furtively obtained news more than its truth value. It hurts India.



A woman seeks permission from a security person standing guard in a street during restrictions in Srinagar

Walk in the park

A visibly mellow Tiger Woods saunters to his all-time-record 82nd PGA Tour win

OVER THE TOP

Meraj Shah



THE PLAYERS SEEMED to be having fun. Smiling after dunking shots into the water, ribbing each other in a scripted sort of way, and just having a marvellously funny day out in the sun in spite of playing spectacularly bad golf. I think it seemed contrived simply because it's hard to believe that any golfer, let alone four of the top players on the planet, can display such a sunny demeanour while their games unravel. Or perhaps, like Sam Snead once said, pro golfers are just so much better at picking themselves up after getting their teeth knocked in every now and then.

That is not to say that *Golf TV's* hyped Skins game featuring Tiger Woods, Jason Day, Rory McIlroy and Hideki Matsuyama that was held at the Accordia Golf Narashino Golf Club in Japan wasn't fun to watch. If the objective of the exhibition-style, made-for-television event, was to be different from the week-after-week fare that we watch on pro tours, then it cer-

tainly nailed that. How often are you likely to see the likes of McIlroy, Matsuyama and Day hit it into the water on the same hole in quick succession? And then proceed to laugh it off? That's precisely what transpired on the 177-yard par-3 fifth hole: all Woods needed to do, (after hitting a reasonably terrible tee shot himself that just about stayed dry) was to get up and down for par to win the hole.

To be fair, we golf fans are just so conditioned to watching hyper-competitive tour events, that the casual attitude of the players in the Skins game seemed a bit calous. But once you got past that, it was fun to watch the players letting their hair down, not dialled in as if their lives depended on it. The 'mystery challenges,' ranged from banal to interesting: the highlight was the par-5 13th hole which each player had to play with just a single club. Day pulled out a 6-iron and hit the shot of the day from a greenside bunker for a tap-in par.

The Skins format is incredibly fun to watch, and justified that when the match went down to the wire—Day sealed it on the 18th hole. The secondary objective of the event was to drum up interest in the PGA Tour's latest event, and it's first ever in Japan—the Zozo Championship—that



Tiger Woods holds a winning trophy at the Zozo Championship, a PGA Tour event, at Narashino Country Club in Inzai, Japan

REUTERS

was played last week. All eyes were on Woods, who returned to competitive golf after yet another knee surgery in August this year, while he wasn't bouncing about, and looked more than a tad tired at the end of the day, Woods appeared to be swinging without discomfort.

While he lost the Japan Skins to Day, it's now apparent that underneath all the banter, Woods used the event to get a measure of the greens, and fine-tune his course strategy. Just how well he did that, was apparent when, in spite of opening with three bogies in the first round of the Zozo Championship, Woods shot nine birdies to open his account with a six-under 64 to take the lead on the first day and then went out and shot an encore the second day. Another seven strokes over the weekend and Woods came home with three shots to spare over local favourite Hideki Matsuyama.

At the turn of the millennium, if you'd ask this columnist, whether Woods would be able to match Sam Snead's mark of career 82 wins on the PGA Tour then I would have told you that it was a no-brainer. At that point Woods seemed set to overhaul pretty much every record in the game. But the script changed after stress fractures in his knee forced Woods out of the game in 2014—for three years he couldn't even swing a club. "Probably the low point was not knowing if I'd ever be able to live pain-free again," he said recently. "Am I going to be able to sit, stand, walk, lay down without feeling the pain

that I was in? I just didn't want to live that way. Is this how the rest of my life is going to be?"

Woods came back after a radical fusion surgery to play his first full season in years and pulled off his 80th win—his first in over five years—at the TOUR Championship in 2018. The high point was winning The Masters Tournament in April this year, his 15th major title and 81st win overall. Unfortunately, his knee gave way yet again, making him a doubtful starter at the Zozo Championship.

No one expected Woods to win; most commentators just wanted to see him get through four rounds without pain, or recurrence of injury. And the man himself looked singularly focused on precisely that—swinging at what seemed like 75%, placing the ball around the course, and making no attempt to overpower the course. What worked like a dream was his trusty blade putter—Woods led the field in strokes gained on the greens and just holed everything. In the process he looked almost nonchalant as he made his way up the leaderboard, sealing it with three shots to spare.

And just like that, he's back, again. Woods is now ranked sixth in the world, and, his body holding up, will headline golf at the Olympics in Japan next year. I wouldn't put it past him to win it. I mean how do you bet against Tiger Woods.

A golfer, Meraj Shah also writes about the game