The international journal of science / 29 June 2023

nature

It's time to talk about the known risks of AI

Forget machine doomsday — what's needed is effective regulation to limit the societal harms artificial intelligence is already causing.

It is unusual to see industry leaders talk about the potential lethality of their own product. It's not something that tobacco or oil executives tend to do, for example. Yet barely a week seems to go by without a tech industry insider trumpeting the existential risks of artificial intelligence (AI).

In March, an open letter signed by Elon Musk and other technologists warned that giant AI systems pose profound risks to humanity. Weeks later, Geoffrey Hinton, a pioneer in developing AI tools, quit his research role at Google, warning of the grave risks posed by the technology. More than 500 business and science leaders, including representatives of OpenAI and Google DeepMind, have put their names to a 23-word statement saying that addressing the risk of human extinction from AI "should be a global priority alongside other societal-scale risks such as pandemics and nuclear war". And on 7 June, the UK government invoked AI's potential existential danger when announcing it would host the first big global AI safety summit this autumn.

The idea that AI could lead to human extinction has been discussed on the fringes of the technology community for years. The excitement about the tool ChatGPT and generative AI has now propelled it into the mainstream. But, like a magician's sleight of hand, it draws attention away from the real issue: the societal harms that AI systems and tools are causing now, or risk causing in future. Governments and regulators in particular should not be distracted by this narrative and must act decisively to curb potential harms. And although their work should be informed by the tech industry, it should not be beholden to the tech agenda.

Many AI researchers and ethicists to whom *Nature* has spoken are frustrated by the doomsday talk dominating debates about AI. It is problematic in at least two ways. First, the spectre of AI as an all-powerful machine fuels competition between nations to develop AI so that they can benefit from and control it. This works to the advantage of tech firms: it encourages investment and weakens arguments for regulating the industry. An actual arms race to produce next-generation AI-powered military technology is already under way, increasing the risk of catastrophic conflict — doomsday, perhaps, but not of the sort much discussed in the dominant 'AI threatens human extinction' narrative.

Second, it allows a homogeneous group of company executives and technologists to dominate the conversation about AI risks and regulation, while other communities are left out. Letters written by tech-industry leaders are "essentially drawing boundaries around who counts as an expert in this conversation", says Amba Kak, director of the AI Now Institute in New York City, which focuses on the social consequences of AI.

> **"AI-powered facial recognition is already being abused by autocratic states to track and oppress people."**

AI systems and tools have many potential benefits, from synthesizing data to assisting with medical diagnoses. But they can also cause well-documented harms, from biased decision-making to the elimination of jobs. AI-powered facial recognition is already being abused by autocratic states to track and oppress people. Biased AI systems could use opaque algorithms to deny people welfare benefits, medical care or asylum — applications of the technology that are likely to most affect those in marginalized communities. Debates on these issues are being starved of oxygen.

One of the biggest concerns surrounding the latest breed of generative AI is its potential to boost misinformation. The technology makes it easier to produce more, and more convincing, fake text, photos and videos that could influence elections, say, or undermine people's ability to trust any information, potentially destabilizing societies. If tech companies are serious about avoiding or reducing these risks, they must put ethics, safety and accountability at the heart of their work. At present, they seem to be reluctant to do so. OpenAI did 'stress-test' GPT4, its latest generative AI model, by prompting it to produce harmful content and then putting safeguards in place. But although the company described what it did, the full details of the testing and the data that the model was trained on were not made public.

Tech firms must formulate industry standards for responsible development of AI systems and tools, and undertake rigorous safety testing before products are released. They should submit data in full to independent regulatory bodies that are able to verify them, much as drug companies must submit clinical-trial data to medical authorities before drugs can go on sale.

For that to happen, governments must establish appropriate legal and regulatory frameworks, as well as applying laws that already exist. Earlier this month, the European Parliament approved the AI Act, which would regulate AI applications in the European Union according to their potential risk — banning police use of live facial-recognition technology in public spaces, for example. There are further hurdles for the bill to clear before it becomes law in EU member states and there are questions about the lack of detail on how it will be enforced, but it could help to set global standards on AI systems. Further consultations about AI risks and regulations, such as the forthcoming UK summit, must invite a diverse list of attendees that includes researchers who study the harms of AI and representatives from communities that have been or are at particular risk of being harmed by the technology.

Researchers must play their part by building a culture of responsible AI from the bottom up. In April, the big machine-learning meeting NeurIPS (Neural Information Processing Systems) announced its adoption of a code of ethics for meeting submissions. This includes an expectation that research involving human participants has been

nature

approved by an ethical or institutional review board (IRB). All researchers and institutions should follow this approach, and also ensure that IRBs — or peer-review panels in cases in which no IRB exists — have the expertise to examine potentially risky AI research. And scientists using large data sets containing data from people must find ways to obtain consent.

Fearmongering narratives about existential risks are not constructive. Serious discussion about actual risks, and action to contain them, are. The sooner humanity establishes its rules of engagement with AI, the sooner we can learn to live in harmony with the technology.

Extreme poverty can be eradicated

To improve millions of lives, find better measures of what constitutes poverty.

By 2030, says the World Bank, something like 574 million people will be living in extreme poverty. That is equivalent to the combined population of the European Union and Japan. The United Nations has a Sustainable Development Goal (SDG) to eradicate extreme poverty by 2030; this was always ambitious, even when policymakers and researchers set the SDGs in 2015. It is now unattainable.

The past few years have bucked a positive trend. Back in 1990, almost two billion people were living under the extreme-poverty line, which the World Bank currently defines as an income of no more than US$2.15 a day at 2017 prices. By 2015, there were fewer than 700 million. Had that trend continued, extreme poverty would have been eliminated by, and possibly before, the SDG target.

But the trend had started to slow by 2020, and the COVID-19 pandemic reversed it, forcing an extra 75 million people below the extreme-poverty line. And the pandemic wasn't the only factor. Soaring food and energy costs after Russia's invasion of Ukraine, ongoing conflicts and, increasingly, the effects of climate change have all played a part. Extreme poverty is starting to decline again, but it will take until 2024 to return to 2019 levels. A rethink in approach is clearly needed — and researchers can get involved.

The World Bank, headquartered in Washington DC, is one of the go-to agencies for both measuring poverty and prescribing solutions to end it. Some 80% of people who escaped poverty between 1993 and 2017 were in China and India — countries that posted impressive economic growth figures for that period. The bank says that, similarly, economic expansion in the countries that now have the highest numbers of people in extreme poverty — most of which are in sub-Saharan Africa — would help them to follow China and India's lead.

> **"By one measure, some 1.2 billion people worldwide are living in acute poverty."**

Some researchers doubt that economic growth automatically leads to reductions in extreme poverty, saying that it often coincides with widening income inequality. But even if we accept the World Bank's premise, economic growth rates across Africa have consistently been much lower than in China and India, and on current trends they will remain so. That poses the question: what other levers can countries pull to improve the lives of hundreds of millions of people?

One answer was established in many now-high-income countries that were rebuilding after the Second World War. A number of countries in Western Europe, for example, established basic social and health-care protections at a time when many nations were dependent on aid from the United States. The principle that these protections help people to escape extreme poverty is just as valid today, and applying it would help countries to build resilience to shocks such as pandemics and climate change.

Counting the cost

Even more fundamentally, researchers are advocating a rethink of how poverty is measured. One problem with using an income-based measure is that it excludes people who are earning more than $2.15 a day but are still unable to fulfil their basic human needs.

In 2010, researchers at the University of Oxford, UK, working with the UN Development Programme (UNDP), created the Multidimensional Poverty Index (MPI; see go.nature.com/3jy2srm). It is an estimate of the number of households facing deprivation when measured by ten basic indicators, including adequate housing, child mortality, clean water, sanitation, cooking facilities and an electricity supply. By this measure, some 1.2 billion people worldwide are living in acute poverty, almost 580 million of whom are in sub-Saharan Africa. The global figure is nearly double that calculated on the basis of income. The UN currently uses the MPI to track progress towards another SDG target: reducing by half the proportion of people experiencing poverty in all its dimensions.

In 2018, inspired by the MPI, the World Bank created the Multidimensional Poverty Measure (MPM; see go.nature.com/3nmhmwh). This assesses the number of households facing deprivation in five dimensions (educational attainment and enrolment, and access to electricity, sanitation and drinking water). But unlike the MPI, the MPM also includes the percentage of households living on less than $2.15 a day.

There are some gaps in the data. Some countries do not provide researchers with access to the relevant data; in others, access is possible but there are few on-the-ground resources to collect the information. But where indicators of multidimensional poverty exist, they provide a nuanced picture and help countries to target interventions.

Researchers who study poverty, and development agencies such as the UNDP, agree that a multidimensional index ought to replace a simpler income-based measure. This September, world leaders will gather in New York City to take stock of the SDGs. One of their tasks must be to continue to nudge the World Bank to make this change happen.