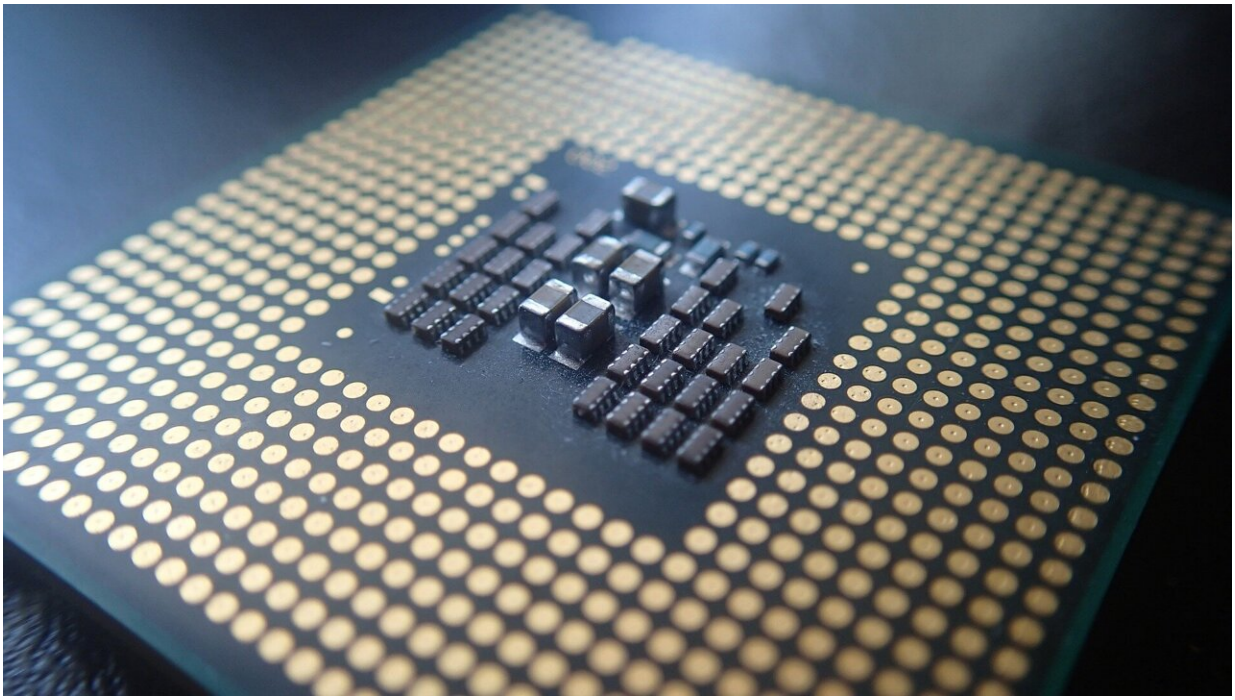
Aim policies at hardware to ensure AI safety, say experts

February 14 2024, by Fred Lewsey



Credit: Pixabay/CC0 Public Domain

A global registry tracking the flow of chips destined for AI supercomputers is one of the policy options highlighted by a major new report calling for regulation of "compute"—the hardware that underpins all AI—to help prevent artificial intelligence misuse and disasters.

Other technical proposals floated by the report include "compute

caps"—built-in limits to the number of chips each AI chip can connect with—and distributing a "start switch" for AI training across multiple parties to allow for a digital veto of risky AI before it feeds on data.

Researchers argue that AI chips and datacenters offer more effective targets for scrutiny and AI safety governance, as these assets have to be physically possessed, whereas the other elements of the "AI triad"—data and algorithms—can, in theory, be endlessly duplicated and disseminated.

The experts point out that powerful computing chips required to drive generative AI models are constructed via highly concentrated supply chains, dominated by just a handful of companies—making the hardware itself a strong intervention point for risk-reducing AI policies.

The report is authored by nineteen experts and co-led by three University of Cambridge institutes—the Leverhulme Center for the Future of Intelligence (LCFI), the Center for the Study of Existential Risk (CSER) and the Bennett Institute for Public Policy—along with OpenAI and the Center for the Governance of AI.

"Artificial intelligence has made startling progress in the last decade, much of which has been enabled by the sharp increase in computing power applied to training algorithms," said Haydn Belfield, a co-lead author of the report from Cambridge's LCFI.

"Governments are rightly concerned about the potential consequences of AI, and looking at how to regulate the technology, but data and algorithms are intangible and difficult to control.

"AI supercomputers consist of tens of thousands of networked AI chips hosted in giant data centers often the size of several football fields, consuming dozens of megawatts of power," said Belfield.

"Computing hardware is visible, quantifiable, and its physical nature means restrictions can be imposed in a way that might soon be nearly impossible with more virtual elements of AI."

The computing power behind AI has grown exponentially since the "deep learning era" kicked off in earnest, with the amount of "compute" used to train the largest AI models doubling around every six months since 2010. The biggest AI models now use 350 million times more compute than thirteen years ago.

Government efforts across the world over the past year—including the US Executive Order on AI, EU AI Act, China's Generative AI Regulation, and the UK's AI Safety Institute—have begun to focus on compute when considering AI governance.

Outside of China, the cloud compute market is dominated by three companies, termed "hyperscalers": Amazon, Microsoft, and Google.

"Monitoring the hardware would greatly help competition authorities in keeping in check the market power of the biggest tech companies, and so opening the space for more innovation and new entrants," said co-author Prof Diane Coyle from Cambridge's Bennett Institute.

The report provides "sketches" of possible directions for compute governance, highlighting the analogy between AI training and uranium enrichment.

"International regulation of nuclear supplies focuses on a vital input that has to go through a lengthy, difficult and expensive process," said Belfield. "A focus on compute would allow AI regulation to do the same."

Policy ideas are divided into three camps: increasing the global visibility

of AI computing; allocating compute resources for the greatest benefit to society; enforcing restrictions on computing power.

For example, a regularly-audited international AI chip registry requiring chip producers, sellers, and resellers to report all transfers would provide precise information on the amount of compute possessed by nations and corporations at any one time.

The report even suggests a unique identifier could be added to each chip to prevent industrial espionage and "chip smuggling."

"Governments already track many economic transactions, so it makes sense to increase monitoring of a commodity as rare and powerful as an advanced AI chip," said Belfield. However, the team point out that such approaches could lead to a black market in untraceable "ghost chips."

Other suggestions to increase visibility—and accountability—include reporting of large-scale AI training by cloud computing providers, and privacy-preserving "workload monitoring" to help prevent an arms race if massive compute investments are made without enough transparency.

"Users of compute will engage in a mixture of beneficial, benign and harmful activities, and determined groups will find ways to circumvent restrictions," said Belfield.

"Regulators will need to create checks and balances that thwart malicious or misguided uses of AI computing."

These might include physical limits on chip-to-chip networking, or cryptographic technology that allows for remote disabling of AI chips in extreme circumstances.

One suggested approach would require the consent of multiple parties to

unlock AI compute for particularly risky training runs, a mechanism familiar from nuclear weapons.

AI risk mitigation policies might see compute prioritized for research most likely to benefit society—from green energy to health and education. This could even take the form of major international AI "megaprojects" that tackle global issues by pooling compute resources.

The report's authors are clear that their policy suggestions are "exploratory" rather than fully fledged proposals and that they all carry potential downsides, from risks of proprietary data leaks to negative economic impacts and the hampering of positive AI development.

They offer five considerations for regulating AI through compute, including the exclusion of small-scale and non-AI computing, regular revisiting of compute thresholds, and a focus on privacy preservation.

Added Belfield, "Trying to govern AI models as they are deployed could prove futile, like chasing shadows. Those seeking to establish AI regulation should look upstream to compute, the source of the power driving the AI revolution.

"If compute remains ungoverned it poses severe risks to society."

More information: Computing Power and the Governance of Artificial Intelligence. www.cser.ac.uk/media/uploads/f … Governance-of-AI.pdf

Provided by University of Cambridge