

Indians training robots who may take over



NIKHIL NARENDRA

Partner, Trilegal

Localisation alone would merely push companies to collect egocentric data elsewhere. Instead, we should create mechanisms to put this data to work for India

FOR OVER THREE decades, every story about India's economic future has rested on the idea of a demographic dividend. We are young, large, and cheaper to employ than other alternatives, and that combination was promised to pull us through a manufacturing-led ascent, as it did for Japan, Korea, Taiwan, and China. It is, however, also becoming obsolete in the Fifth Industrial Revolution.

In factories across Tirupur, NCR, Coimbatore, and Surat, Indian workers are being fitted with head-mounted cameras and/or smart glasses to capture egocentric data. The footage they generate of folding, sorting, packing, assembling, soldering, and sewing is being shipped to companies that aggregate it into data sets and sell it to robotics companies for developing humanoid robots.

Many of these services offer a productivity analytics report at minimal or no cost as an inducement to allow recording. In most cases, the worker has neither been asked nor told what the footage is for. Even when it is provided at a cost, the final metadata and the learnings belong to the company collecting the egocentric data.

One's instinct would be to ask if this is legal under the Digital Personal Data Protection Act. It mostly is. Once faces are blurred, what remains are hands and motion, and that data falls outside the category of personal data protected by the DPDPA. The real problem is not worker privacy, but the impact on India's economy.

The egocentric videos collected from

Indian factories serve as training data for the next generation of physical AI. Humanoid robots, trained on hundreds of millions of hours of human motion, are projected to enter industrial-scale deployment within the next decade.

When ready, these robots are likely to be deployed in countries where these companies are domiciled and have better infrastructure, like the US, Germany, or China. Very few of the early units will reach India, as it lacks the capital base, infrastructure, or energy capacity to absorb them at scale. By the time the second wave arrives, the manufacturing we hoped could absorb our working population would have moved to wherever the robots are deployed.

In this reshoring driven by automation, the factories that were supposed to lift India out of agricultural underemployment may never arrive. The harshest part of the design is that Indian workers are training the very machines that will make them jobless, for wages that would hardly feed a family in a tier-II city.

While there has been no Indian policy response to this, even before egocentric data models were a thing, the Economic Survey of 2018-19 devoted an entire chapter to data and described Indians' data as a national resource, a public good to be harnessed for economic benefit. We also had the Kris Gopalakrishnan Committee report on a

Non-Personal Data Governance Framework, 2020. The report's first principle was that Indians have rights over our data. It proposed a framework to address the kind of industrial training data now being exported from the country. The report gathers dust. No institutional architecture was built on it.

Meanwhile, China has built the playbook for non-personal data that feeds into its AI-robotic production line. In April 2020, the State Council formally designated data as a factor of production alongside land, labour, capital, and technology. Important or core industrial data must be processed in China and can only be exported after a security review.

The same data flows through a domestic ecosystem. Beijing, Shanghai, Shenzhen, Guangzhou, and Guiyang host state-backed data exchanges where industrial data sets are registered, priced, and traded under enforceable use restrictions for Chinese firms to create robots. The state is simultaneously the largest buyer of locally built humanoids. The data, the firms, the buyers, and the factories are all Chinese.

This is not to say that India needs to copy China. However, it's time that we act quickly, re-examine, and revive the discussion around non-personal data, including industrial data.

In India, discussions of data

sovereignty are often conflated with localisation. Localisation alone would simply push these companies to collect egocentric data elsewhere, in Kenya, Ethiopia, or Bangladesh. Instead, we should create mechanisms to put this data to work for India, enabling it to manufacture robots locally.

For this, a national industrial data exchange could be considered that registers Indian-origin training data, with contributors represented as a beneficiary class. It will enable the same data to be used for trading and other purposes within India, promote local R&D and manufacturing, and have revenues support local skilling.

The government also needs to play an anchor role in building domestic demand for Indian-built robots. We already have some good examples of the government promoting robotics. Bandicoot, a manhole-cleaning robot, has already been deployed across India to stop manual scavenging. Indian Railways alone, with its vast yards and tracks still cleaned largely by humans, could anchor a similar industry.

None of this is protectionism, but a policy that recognises the demographic dividend is not a permanent advantage. We have perhaps 10 years before the wave of reshoring. If we do not act, we will discover that our promised future was exported, packaged, and sold back to us as a finished good; the same extractive pattern during the British era, only now the raw material is our labour data, not our cotton or spices.

The government also needs to play an anchor role in building domestic demand for Indian-built robots