

Business Standard

AI's fuel: Why the world is teaming up for datasets

Data fed into an AI algorithm determines its efficiency. The right dataset prevents inbuilt biases and errors. Studies estimate the market for training datasets is worth over \$8 bn

Pranjal Sharma |

In any plan for a digital revolution, reskilling people is cited as an important requirement. Similarly, in a technology-driven world, the right data is needed for building artificial intelligence (AI) algorithms. As a well-skilled professional creates efficient output, strong training data can build an effective AI algorithm.

Training data is the fuel for AI. Data fed into an AI algorithm is a key determinant of its efficiency and quality of output. "Data training refers to providing a machine learning algorithm with labelled or categorised examples to assist it in recognising patterns. These examples can range from images and text documents to numerical values. The goal is for the algorithm to create a representation based on input-output

correlations, allowing it to generate accurate predictions when exposed to new, unseen data,” according to IEEE Computer Society.

Training and testing in the field of machine learning for AI is an important part of enabling accuracy and efficacy. The right type of dataset prevents inbuilt biases and errors. AI algorithms for predictive medicine will need training data on various kinds of patients. The training dataset should have the relevant configuration of patients from every part of the world.

Designing a strong dataset for training requires diverse inputs. For this experts are arguing for deeper collaboration in sharing training datasets. Data can be taken from open sources or from public and private sector organisations that have information on consumers and citizens. A global protocol on data collaboration is equally important while ensuring that personal privacy is not undermined.

Consulting firm McKinsey has estimated that \$3 trillion in market value can be unlocked every year globally with data collaboration. Collaboration can be in the form of public-private partnerships or between commercial entities.

“We see a clear potential to unlock significant economic value by applying advanced analytics to both open and proprietary knowledge. Open data can become an instrument for breaking down information gaps across industries, allowing companies to share benchmarks and spread best practices that raise productivity. Blended with proprietary datasets, it can propel innovation and help organisations replace traditional and intuitive decision-making approaches with data-driven ones,” according to McKinsey.

Major companies have created guidelines for data collaboration. Microsoft has said its five principles for data collaboration are open, usable, empowering, secure, and private. IBM Corporation has released a dataset of more than a million images of people to improve facial recognition systems supported by AI.

“Due to the rapid adoption of artificial intelligence technology, the need for training datasets is rising exponentially. To make the technology more versatile and accurate with its predictions, many companies are entering the market by releasing various datasets operating across different use cases to train the machine learning algorithm,” said a report by Grandview Research. Companies like Google, Microsoft, Apple Inc, and Amazon have been focusing on developing various AI training datasets. Amazon has released a dataset of commonsense dialogue to support analysis of open-domain conversation.

The market for training datasets has been estimated to be worth more than \$8 billion by various studies. Nasscom, which represents India's technology industry, has estimated that a comprehensive data collaboration and utilisation plan can add \$500 billion to India's gross domestic product by 2025.

Apart from private companies, collaboration can also occur between governments. It is the next phase of multilateral cooperation where countries collaborate in exchanging datasets for objectives like climate change, public health and urban planning. India and other countries are setting up structures to collaborate on datasets. India has set up AI and data collaborations with the US and Germany. Collaboration in training datasets will be multidimensional and it will be determined by the needs of an economy or a sector. Much like the training professionals.

First Published: Nov 26 2023 | 10:15 PM IST

Page URL :https://www.business-standard.com/technology/tech-news/ai-s-fuel-why-the-world-is-teaming-up-for-datasets-123112600913_1.html