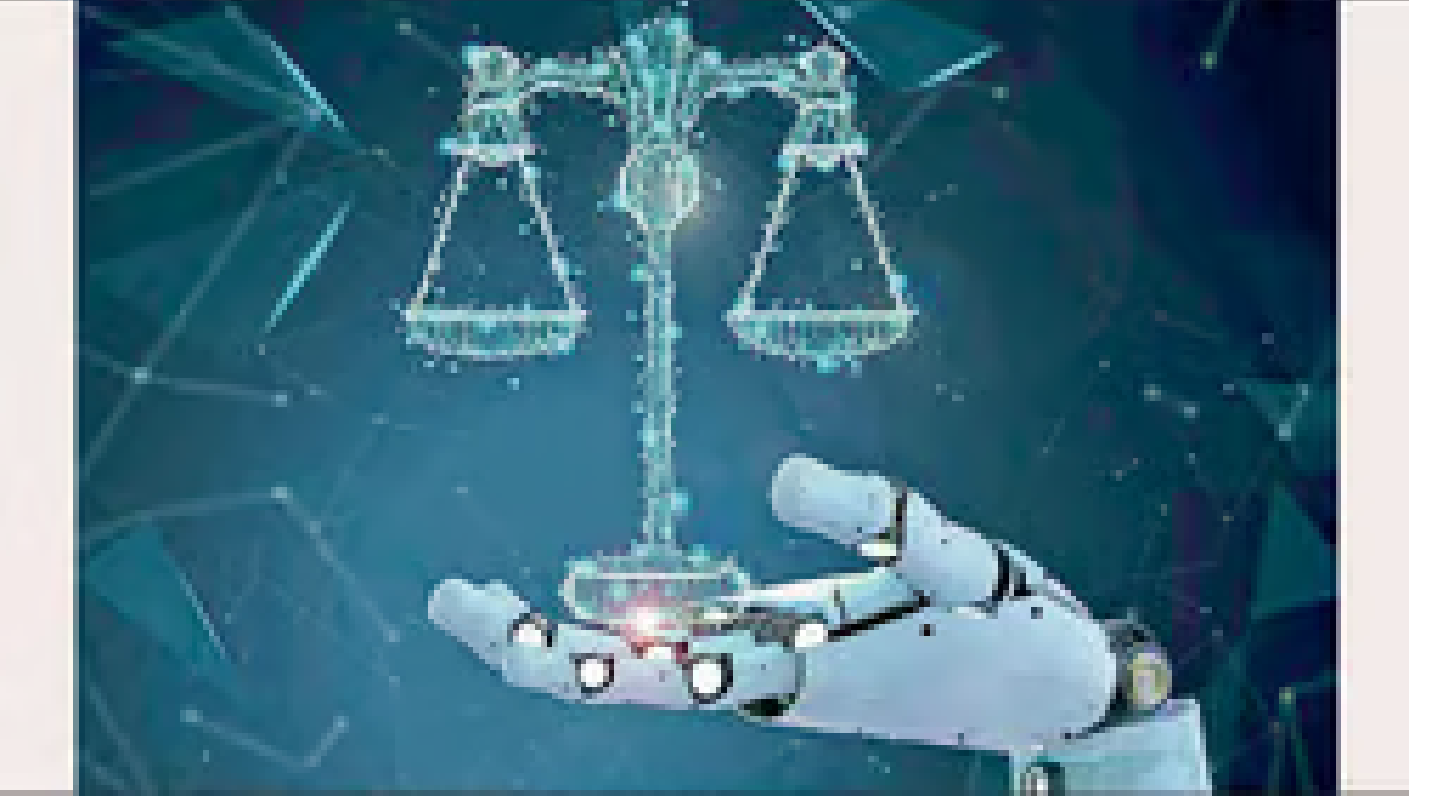


संपादकीय

भाष्य : सत्य-असत्यासी 'तंत्र' केले ग्वाही

पॅरिसमध्ये पार पडलेल्या व 'आर्थिक सहकार्य आणि विकास संघटने'ने आयोजित केलेल्या एका कार्यक्रमात जपानचे पंतप्रधान किशिदा यांनी एआय नियमन आराखड्याची घोषणा केली.



AI Regulatory Frameworksaka

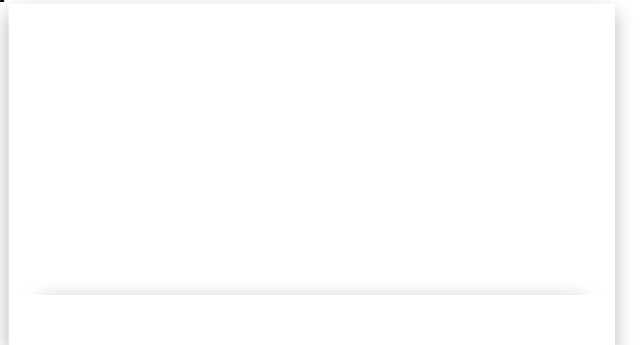
सकाळ वृत्तसेवा

Updated on: 07 May 2024, 6:25 am



- डॉ. योगेश कुलकर्णी

पॅरिसमध्ये पार पडलेल्या व 'आर्थिक सहकार्य आणि विकास संघटने'ने आयोजित केलेल्या एका कार्यक्रमात जपानचे पंतप्रधान किशिदा यांनी एआय नियमन आराखड्याची घोषणा केली. हे प्रयत्न आवश्यक आहेत, यात शंका नाही. मात्र नियमांची चौकट व आखलेली लक्ष्मणरेखा ही काळानुसार बदलणा



'डीप फेक व्हिडिओ'च्या भयानक परिणामांचा फटका खुद्द पंतप्रधान नरेंद्र मोदी आणि केंद्रीय गृहमंत्री अमित शहा यांना बसल्यानंतर आपल्याकडे या प्रकारांना वेसण कशी घालायची याची चर्चा सुरू झाली. परंतु जगभरच या नियमनाचे स्वरूप कसे असावे, यावर ऊहापोह सुरू झाला आहे.

ऑस्ट्रेलियातील क्वीन्सलँड विद्यापीठाने कृत्रिम बुद्धिमत्तेविषयी (आर्टिफिशिअल इंटेलिजन्स, एआय) केलेल्या सर्वेक्षणात असे आढळून आले की, जवळजवळ दोन तृतीयांश लोकांना वाटते की, 'एआय'वर जरूर नियंत्रण असावे. पॅरिसमध्ये पार पडलेल्या व 'आर्थिक सहकार्य आणि विकास संघटने'ने आयोजित केलेल्या एका कार्यक्रमात जपानचे पंतप्रधान किशिदा यांनी एआय नियमन आराखड्याची घोषणा केली. त्यात 'कृत्रिम बुद्धिमत्तेने (जनरेटिव्ह एआय, जेन-एआय) उभी केलेली आव्हाने, त्याचा जागतिक प्रभाव आणि त्याच्या नियमन-नियंत्रणाविषयीच्या मुद्द्यांचा ऊहापोह केला आहे.

'जेन-एआय' ही 'एआय'चीच एक उपशाखा असून त्यात सर्जनावर म्हणजेच नवनवीन गोष्टी (चित्रे, आवाज, भाषा, इ.) बनवणाऱ्या प्रणालींचा समावेश होतो. चॅटजीपीटी हे त्याचे प्रचलित उदाहरण आहे, ज्यात तुम्ही प्रामुख्याने विविध भाषाविषयक गोष्टींची निर्मिती करू शकता. सूचना (प्रॉम्प्ट) देईल, तसे उत्तर ही प्रणाली देते.

कथा लिहायला सांगा, कविता लिहायला सांगा, संगणक प्रोग्रॅम लिहायला सांगा, सारांश काढायला सांगा, इत्यादी अनेक गोष्टी ते लीलया करीत असल्याने जगभरात ते कमालीचे लोकप्रिय झालेले आहे. त्याच पद्धतीने, प्रॉम्प्ट देऊन चित्रे बनवणाऱ्या, ध्वनी तयार करणाऱ्या प्रणालीही आता उपलब्ध आहेत. पण त्यामुळे अशी परिस्थिती निर्माण झाली आहे की, एखादा लेख अथवा चित्र माणसाने बनवले आहे की जेन-एआयने, हे समजणे अवघड होत चालले आहे.

त्याचा 'डीप फेक' (उदाहरणार्थ, एखाद्या व्हिडीओमधील व्यक्तीचा फक्त चेहरा बदलून दुसऱ्या-प्रसिद्ध व्यक्तीचा चेहरा त्यावर लावणे, जेणेकरून असा भास होईल, की त्या प्रसिद्ध व्यक्तीचाच तो व्हिडिओ आहे.) अगदी सारख्या, बनावट आवाज वापरण्यासारख्या, खोटी माहिती पसरवण्यासारख्या गोष्टींमध्ये ही या तंत्राचा वापर सुरू झालेला आहे. त्यावर आळा कसा घालता येईल, यासंदर्भात जनतेत, कंपन्यांमध्ये आणि सरकार-दरबारी चर्चा-योजना सुरू झाल्या आहेत.

आंतरराष्ट्रीय आराखडा जपानने मागील वर्षी 'हिरोशिमा एआय प्रोसेस'ची घोषणा केली होती. ज्या ज्या कंपन्या एआय प्रणाली-प्रारूपे (मॉडेल्स) बनवत आहेत, त्यांच्यासाठी आचारसंहिता आणि मार्गदर्शक तत्त्वे लागू करण्याची मागणी त्यात करण्यात आली होती. या आराखड्यास ४९ देशांनी मान्यता देऊन तो स्वीकारला होता. आता त्याचा पुढील टप्पा म्हणून पंतप्रधान किशिदा यांच्याकडून हा आंतरराष्ट्रीय आराखडा जाहीर करण्यात आला आहे. यातील तरतुदी लवकरच प्रसिद्ध केल्या जाणार आहेत.

एआय नियमनाचे काम एकट्या-दुकट्या देशाने करून भागणार नाहीये. जागतिक सहयोग-सहकार्याची गरज असणार आहे. समान नियंत्रक तत्त्व झाल्यावर त्यांच्या अंमलबजावणीस सुलभता येईल. हे नियमन कशा प्रकारचे त्यामुळे काय फायदे होतील याविषयी काही मुद्दे पाहू.

'जेन-एआय'चा वापर संगणक प्रोग्रामिंग, वैद्यकीय, विधी, वित्त यांसारख्या सर्जनशील क्षेत्रात पण जोरात होतो आहे. त्यामुळे तेथे अनेक प्रकारचे धोके मुळात हे नियमन शक्य आहे का? आणि असेल तर कसे?

इकडे आड, तिकडे विहीर

जेन-एआय च्या प्रणाल्या व प्रारूपे अजस्त असतात. त्यांच्या प्रशिक्षणात करोडोनी शब्दसंग्रहांचा उपयोग केलेला असतो. त्यामुळे त्याने दिलेले उत्तर हे नक्की कोठली माहिती वापरून तयार केले आहे, हे समजणे दुरापास्त असते. याचाच अर्थ, एखादे चुकीचे उत्तर आले किंवा खोटी माहिती आली तर ती कशामुळे आली हे शोधणे जिकीरीचे ठरते.

मग नियंत्रण ठेवायचे कसे? दररोज नव-नवीन मॉडेल येत आहेत, त्यांचा प्रशिक्षणाचा डेटा वेगळा, त्यांचे न्यूरल नेटवर्क (प्रारूपाचा संगणकीय आराखडा) वेगळे. या सर्वांवर कसे लक्ष ठेवायचे, हे आव्हान आहे. सर्व तपासत बसलो तर त्यातच वेळ जाणार, नवीन प्रणालींना यायला वेळ लागणार आणि त्यामुळे पुढील संशोधनाला खीळ बसणार, असा पेच आहे. इकडे आड तर तिकडे विहीर. यावर उपाय शोधण्यासाठी काही मार्गदर्शक तत्त्वे सांगता येतील.

पहिले म्हणजे 'स्वामित्व'. ज्यांनी एखादी प्रणाली बनवली असेल, त्यांनी त्याच्यातून येणाऱ्या उत्तरांची पूर्ण जबाबदारी घेणे. 'आम्ही प्रशिक्षणासाठी दुसऱ्यांचा डेटा घेतला आहे' यासारख्या सबबी चालणार नाहीत. उत्तरे देण्याआधी त्यांची नैतिकतेची, सत्यतेची तपासणी आणि संगणकीय कुंपणे (गार्डरेल्स) लावण्याची जबाबदारी त्यांची. एखाद्या उत्तराने कोणाची मानहानी झाली तर न्यायालयात तोंड देण्याची जबाबदारी पण त्यांची.

दुसरे तत्त्व पारदर्शकतेचे. ज्यांच्या 'जेन-एआय' प्रणाली-प्रारूपे 'बंद-गुप्त' (क्लोझ्ड सोर्स) स्वरूपाची आहेत, त्यांवर शंका जास्त येते. यावर उपाय म्हणून काही कंपन्यांनी त्यांची प्रारूपे खुली (ओपन-सोर्स) केली आहेत. त्यामुळे वापरलेले न्यूरल नेटवर्क काय आहे आणि (उघड केले असेल तर) कोठला डेटा प्रशिक्षणाला वापरला आहे, याची माहिती सार्वजनिक केलेली असते.

साधारणतः त्या सर्वसामान्यांसाठी वापरायला फुकट सुद्धा असतात. यामुळे चांगली तपासणी होऊ शकते, बदल सुचवले जाऊ शकतात. स्वामित्वाची जबाबदारी कमी होत नसली तरी लोकांना त्याविषयी जास्त विश्वास वाटू शकतो. काही शास्त्रज्ञ ओपन-सोर्स-मॉडेल्स याची भलावण यासाठीच करतात.

पण अशी प्रारूपे सर्व कंपन्या लोकांना फुकट का वापरायला देतील? त्यांच्या प्रशिक्षणासाठी खर्च केलेले कोट्यवधी डॉलर मग त्या परत कसे मिळवणार? ते धंद्याला मारकच नाही का? तर यावर उपाय काय, हा कळीचा मुद्दा आहे. तिसरे तत्त्व म्हणजे 'डेटा'.

जेन-आय मॉडेल्सच्या प्रशिक्षणासाठी वापरण्यात आलेला डेटा कोठून आणला आहे? त्याच्या मालकांची परवानगी घेतली आहे का? त्यात काही पक्षपाती तरतुदी दडल्या आहेत का? त्यात कोणाची खासगी अथवा गोपनीय माहिती आहे का? असे असंख्य प्रश्न निर्माण होतात. 'जसा डेटा तसे मॉडेल' हे ब्रीद असल्याने डेटा-तपासणीचा मुद्दा नियामकांना विचारात घ्यावा लागेल.

चौथे तत्त्व 'अंमलबजावणी' विषयी. मार्गदर्शक तत्त्वे बनवली तरी ती कशी वापरणार, प्रत्येक नवीन मॉडेलची तपासणी करून ते योग्य-सुरक्षित असल्याचे प्रमाणित करणार का, डेटा ऑडिटिंग कसे करणार, त्याच्यासाठी लागणाऱ्या संगणकीय प्रणाली कोण बनवणार, आणि हे करण्यासाठी प्रशिक्षित मनुष्यबळ कसे निर्माण करणार, असे प्रश्न आहेत. प्रारूपे ओपन-सोर्स असतील तर या संदर्भात जनसामान्यांची मोठी मदत होऊ शकते.

अशा मुद्द्यांचा सखोल विचार करून, जागतिक सहकार्याने, तज्ज्ञांची मते जाणून घेऊन नियामक आराखडा बनवावा लागणार आहे. 'एआय'वर नियंत्रण तर नक्कीच असावे; पण त्यासाठी केलेली नियमांची चौकट व लक्ष्मणरेखा ही काळानुसार बदलणारी, नूतन तंत्रज्ञानानुसार लवचिक होणारी आणि सुधारणेस वाव असणारी असावी, ही अपेक्षा. थोडक्यात, सत्य-असत्यासी 'तंत्र' केले ग्वाही...असे म्हणावेसे वाटते. अर्थात तुकोबांची क्षमा मागून.

- लेखक 'कृत्रिम बुद्धिमत्ता'(एआय) विषयाचे सल्लागार आहेत.yogeshkulkarni@yahoo.com

Read latest Marathi news, Watch Live Streaming on Esakal and Maharashtra News. Breaking news from India, Pune, Mumbai. Get the Politics, Entertainment, Sports, Lifestyle. Jobs. and Education updates. And Live taja batmya on Esakal Mobile App. Download the Esaka IOS.

