



Yogesh Haribhau Kulkarni • You

AI Advisor (Helping people/organizations in their AI j...
now • 🌐

🎉 My article on formulating a framework for regulating #ArtificialIntelligence (#AI) has been published in the local Marathi newspaper, Sakal, today! 📰

In my opinion, the key pillars for effective AI regulations are:

- 🔑 Ownership
- 💡 Transparency
- 📊 Data
- ⚖️ Enforcement

Stay tuned as I'll be posting a Medium blog of the same content, in the comments below, for easier reading. 📄✍️

#AIRegulations #AIGovernance #TechPolicy #FutureofAI
#DataPrivacy #EthicalAI #Pune #sakal #artificialintelligence
#regulations #India #data #security

Sakal Media Group Niranjan Agashe Abhijit Pawar Dr. Vinay
Sahasrabuddhe Kaushal Inamdar

सत्य-असत्यासी 'तंत्र' केले ग्वाही

भाष्य



डॉ. योगेश कुलकर्णी

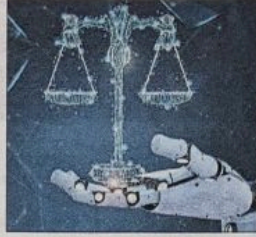
पॅरिसमध्ये पार पडलेल्या व 'आर्थिक सहकार्य आणि विकास संघटने'ने आयोजित केलेल्या एका कार्यक्रमात जपानचे पंतप्रधान किशिदा यांनी एआय नियमन आराखड्याची घोषणा केली. हे प्रयत्न आवश्यक आहेत, यात शंका नाही. मात्र नियमांची चौकट व आखलेली लक्षमणरेखा ही काळानुसार बदलणारी, नूतन तंत्रज्ञानानुसार लवचिक होणारी असावी.

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

'डी प फेक बिडिओ'च्या भयानक परिणामांचा फटका खूब पंतप्रधान नरेंद्र मोदी आणि केंद्रीय गृहमंत्री अमित शहा यांना बसल्यानेतर आपल्याकडे या प्रकारांना वेगाने कसा घालण्याची याची चर्चा सुरू झाली. परंतु जगभरच या नियमनाचे स्वरूप कसे असावे, यावर उहापोह सुरू झाला आहे. ऑस्ट्रेलियातील क्वीन्सलँड विद्यापीठाने कृत्रिम बुद्धिमत्तेविषयी (आर्टिफिशियल 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 बुद्धिमत्ते' (क्वॉटिफ एआय, जेन-एआय) उभी केलेली आम्हाला, त्याचा जागतिक प्रभाव आणि त्याच्या नियमन-नियंत्रणाविषयीच्या मुद्द्यांचा उहापोह केला आहे.

'जेन-एआय' ही 'एआय'चीच एक उपशाखा असून त्यात सर्जनात्मक म्हणजेच नवनवीन गोष्टी (चिन्हे, आविष्कार, 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 लोळ्या करीत असल्याने जगभरात ते कमालीचे लोकप्रिय झालेले आहे. त्याच पद्धतीने, प्रॉम्प्ट देऊन चिन्हे बनवणाऱ्या, ध्वनी तयार करणाऱ्या प्रणालीही आता उपलब्ध आहेत. पण त्यामुळे अशी परिस्थिती निर्माण झाली आहे की, एखादा लेख अथवा चित्र मागसने बनवले आहे की जेन-एआयचे, हे समजणे अवघड होत चालले आहे. त्याचा 'डीप फेक' (उदाहरणार्थ, एखाद्या बिडिओमधील व्यक्तीचा फक्त चेहरा बदलून दुसऱ्या-प्रसिद्ध व्यक्तीचा चेहरा त्यावर लावणे, जेणेकरून असा भास होईल, की त्या प्रसिद्ध व्यक्तीचाच तो बिडिओ आहे.) अगदी सारख्या, बनावट आवारा वापरण्यासारख्या, खोटी माहिती पसरवण्यासारख्या गोष्टींमध्ये ही या तंत्राचा वापर सुरू झालेला आहे. त्यावर आज कासा घालता येईल, यासंदर्भात जनेतेत, कंपन्यांमध्ये आणि सरकार-दरबारी चर्चा-योजना सुरू झाल्या आहेत.

जपानने मागील वर्षी 'हिरोशिमा एआय प्रॉसेस'ची घोषणा केली होती. ज्या ज्या कंपन्या एआय प्रणाली-प्राप्ते (मॉडिफ्स) बनवत आहेत, त्यांच्यासाठी आचारसंहिता आणि मार्गदर्शक तत्वे 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 सततुदी लवकरच प्रसिद्ध केल्या जाणार आहेत. एआय नियमनाचे काम एकट्या-दुकट्या देशाने करून भागणार नाहीये. जागतिकीकरणामुळे आणि परस्पर अवलंबित्वामुळे जागतिक सहयोग-सहकार्याची गरज असणार आहे. समान नियंत्रक तत्वांमुळे सुसूत्रता तर येईलच पण नियमन-मानके तयार झाल्यावर त्यांच्या अंमलबजावणीस सुलभता येईल. हे नियमन कशा प्रकारे केले जाईल, त्यात नक्की काय पहिले जाईल आणि त्यामुळे काय फायदे होतील याविषयी काही मुद्दे पार. 'जेन-एआय'चा वापर संगणक प्रोग्रॅमिंग, वैद्यकीय, विधी, वित्त यांसारख्या क्षेत्रातच नाही तर साहित्य-कला-चित्रपट यांसारख्या सर्जनशील क्षेत्रात पण जोरात होतो आहे. त्यामुळे तेथे अनेक प्रकारचे धोके निर्माण होत आहेत. तर यांचे नियमन कसे करायचे? मुळात हे नियमन शक्य आहे का? आणि असेल तर कसे?

इकडे आड, तिकडे विहीर

जेन-एआय च्या प्रणाल्या व प्राप्ते अजूनच असतात. त्यांच्या प्रशिक्षणात करोडोंनी 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 विकिरीचे ठरते. माग नियंत्रण ठेवायचे कसे? दररोज नव-नवीन मॉडेल येत आहेत, त्यांचा प्रशिक्षणाचा डेटा वेगळा, त्यांचे न्यूरल नेटवर्क (प्राक्पाचा 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 खोड बसणार. असा पेच आहे. इकडे आड तर तिकडे विहीर. यावर उपाय शोधण्यासाठी काही मार्गदर्शक तत्वे सांगता येतील.

पहिले म्हणजे 'स्वामित्व'. ज्यांनी एखादी प्रणाली

बनवली असेल, त्यांनी त्याच्यातून येणाऱ्या उत्तरांची पूर्ण जबाबदारी घेणे. 'आम्ही प्रशिक्षणासाठी दुसऱ्यांचा डेटा घेतला आहे' यांसारख्या सबबी चालणार नाहीत. उत्तरे देण्याआधी त्यांची नैतिकतेची, सत्यतेची तपासणी आणि संगणकीय कुणपे (गॉडीटल्स) लावण्याची जबाबदारी त्यांची. एखाद्या उत्तराने कोणाची मानही झाली तर न्यायालयात तोंड देण्याची जबाबदारी पण त्यांची. दुसरे तत्त्व पारदर्शकतेचे. ज्यांच्या 'जेन-एआय' प्रणाली-प्राप्ते 'बंद-गुप्त' (क्लोस्ड सोर्स) स्वरूपाची आहेत, त्यावर शंका जास्त येते. यावर उपाय म्हणून काही कंपन्यांनी त्यांची प्राप्ते खुली (ओपन-सोर्स) केली आहेत. त्यामुळे वापरलेले न्यूरल नेटवर्क काय आहे आणि (उघड केले असेल तर) कोठल्या डेटा प्रशिक्षणात वापरला आहे, याची माहिती सार्वजनिक केलेली असेल. साधारणतः त्या सर्वसामान्यांसाठी वापरण्याला फुकट सुद्धा असतल. यामुळे चांगली तपासणी होऊ शकते, बदल सुचकले जाऊ शकतात. स्वामित्वाची जबाबदारी कमी होत नसली तरी लोकांना त्याविषयी जास्त विश्वास वाढू शकतो. काही शास्त्रज्ञ ओपन-सोर्स मॉडेलस याची भलायका यमासाठीच करतात. पण अशी प्राप्ते सर्व कंपन्या लोकांना फुकट का वापरण्यास देतील? त्यांच्या प्रशिक्षणासाठी खर्च केलेले कोट्यवधी डॉलर माग त्या परत कसे मिळवणार? ते घडाला मारकच नाही का? तर यावर उपाय काय, हा कळीचा मुद्दा आहे. तिसरे तत्त्व म्हणजे 'डेटा'. जेन-एआय मॉडेलसच्या प्रशिक्षणासाठी वापरण्यात आलेला डेटा कोठून आणला आहे? त्याच्या मालकांची परवानगी घेतली आहे का? त्यात काही पक्षपाती सततुदी दडल्या आहेत का? त्यात कोणाची खासगी अवका गोपनीय माहिती आहे का? असे असंख्य प्रस्न निर्माण होतात. 'जसा डेटा तसे मॉडेल' हे ब्रीद असल्याने डेटा-तपासणीचा मुद्दा नियामकांना विचारात घ्यावा लागेल.

चौथे तत्त्व 'अंमलबजावणी' विषयी. मार्गदर्शक तत्वे बनवली तरी ती कशी वापरणार, प्रत्येक नवीन मॉडेलची तपासणी करून ते योग्य-सुरक्षित असल्याचे प्रमाणित करणार का, डेटा ऑडिटिंग कसे करणार, त्यांच्यासाठी लागणाऱ्या संगणकीय प्रणाली कोण बनवणार, आणि हे करण्यासाठी प्रशिक्षित मनुष्यबळ कसे निर्माण करणार, असे प्रस्न आहेत. प्राप्ते ओपन-सोर्स असतील तर या संदर्भात जनसामान्यांची मोठी मदत होऊ शकते. अशा मुद्द्यांचा सखोल विचार करून, जागतिक सहकार्य, तज्ज्ञांची मते जाणून घेऊन नियामक आराखडा बनवावा लागणार आहे. 'एआय'वर नियंत्रण तर नक्कीच असावे; पण त्यासाठी केलेली नियमांची चौकट व लक्षमणरेखा ही काळानुसार बदलणारी, नूतन तंत्रज्ञानानुसार लवचिक होणारी आणि सुधारणेस वाव असणारी असावी, ही अपेक्षा. थोडक्यात, सत्य-असत्यासी 'तंत्र' केले ग्वाही... असे म्हणजेसे वाटते. अर्थात तुकोबांची सत्ता मरण.