

आखणी लक्ष्मणरेखेची

ऑस्ट्रेलियातील क्वीन्सलँड विद्यापीठाने कृत्रिम बुद्धिमत्ते (आर्टिफिशियल इंटेलिजन्स, एआय) विषयी केलेल्या सर्वेक्षणात असे आढळून आले की जवळ जवळ दोन तृतीयांश लोकांना वाटते की एआय वर जरूर नियंत्रण असावे. एआयचे फायदे सर्वांना माहिती असले तरी त्याच्या संभाव्य धोक्याची जाणीव पण, 'डीप फेक' (बनावटी चित्रे-व्हिडिओ) यांसारख्या उदाहरणांनी झाली आहे. प्रगत देश याविषयी नियमन-कायद्यांविषयी प्रयत्न करीत आहेत. याच संदर्भात, पॅरिसमध्ये पार पडलेल्या व 'आर्थिक सहकार्य आणि विकास संघटने'ने आयोजित केलेल्या एका कार्यक्रमात जपानचे पंतप्रधान किशिदा यांनी एआय नियमन आराखड्याची घोषणा केली. त्यात सृजन कृत्रिम बुद्धिमत्ते (जनरेटिव्ह एआय, जेन-एआय) ने उभी केलेली आव्हाने, त्याचा जगातील प्रभाव आणि त्याच्या नियमन-नियंत्रणाविषयीच्या मुद्द्यांचा उहापोह केला आहे. याविषयी जरा सखोल जाणून घेऊयात.

जेन-एआय ही एआयचीच एक उपशाखा असून त्यात सृजनावर म्हणजेच नव-नवीन गोष्टी (चित्रे, आवाज, भाषा, ई.) बनवणाऱ्या प्रणालींचा समावेश होतो. चॅटजिपीटी 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 प्रणाल्या पण 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 व्हिडीओ मधील व्यक्तीचा फक्त चेहरा बदलून दुसऱ्या-प्रसिद्ध व्यक्तीचा चेहरा त्यावर लावणे, जेणेकरून असा भास होईल कि त्या प्रसिद्ध व्यक्तीचाच तो व्हिडीओ आहे) सारख्या, बनावट आवाज वापरण्यासारख्या, खोटी माहिती पसरवण्यासारख्या गोष्टींमध्ये ही वापर सुरु झालेला आहे. त्यावर आळा कसा घालता येईल यासंदर्भात जनतेत, कंपन्यांमध्ये आणि सरकार-दरबारी चर्चा-योजना सुरु झाल्या आहेत.

जपानने मागील वर्षी 'हिरोशिमा एआय प्रोसेस'ची घोषणा केली होती. ज्या ज्या कंपन्या एआय प्रणाली-प्रारूपे (मॉडेल्स) 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 तरतुदी 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 बघुयात.

जेन-एआयचा वापर संगणक प्रोग्रामिंग, वैद्यकीय, विधी, वित्त यांसारख्या क्षेत्रातच नाही तर साहित्य-कला-चित्रपट यांसारख्या सृजनशील क्षेत्रात पण जोरात होतो आहे. त्यामुळे तेथे अनेक प्रकारचे धोके निर्माण होत आहेत. तर यांचे नियमन कसे करायचे, प्रथमतः नियमन करणे शक्य आहे का? आणि असेल तर कसे?, ते पाहू.

जेन-एआय च्या प्रणाल्या व प्रारूपे अजस्र मोठ्या 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, की एखादे चुकीचे उत्तर आले किंवा खोटी माहिती आली तर ती कशामुळे आली हे शोधणे जिकिरीचे ठरते. मग नियंत्रण ठेवायचे तरी कसे? पुन्हा दररोज नव-नवीन मॉडेल्स येत आहेत, त्यांचा प्रशिक्षणाचा डेटा (विदा, माहिती) वेगळा, त्यांचे न्यूरल नेटवर्क (प्रारूपाचा संगणकीय आराखडा) वेगळे, या सर्वांवर कसे लक्ष

ठेवायचे, हे महत्वाचे आव्हान आहे. सर्व तपासत बसलो तर त्यातच वेळ जाणार, नवीन प्रणाल्यांना यायला वेळ लागणार आणि त्यामुळे पुढील संशोधनाला खीळ बसणार, असा पेच निर्माण झाला आहे. इकडे आड तर तिकडे विहीर. यावर उपाय शोधण्यासाठी काही मार्गदर्शक तत्वांचा विचार मात्र केला जाऊ शकतो.

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

दुसरे तत्व पारदर्शिकतेचे. ज्यांच्या जेन-एआय प्रणाल्या-मॉडेल्स बंद-गुप्त (क्लोज्ड सोर्स) स्वरूपाच्या आहेत त्यांवर जनमानसात शंका जास्त येते. यावर उपाय म्हणून काही कंपन्यांनी त्यांची मॉडेल्स 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 या 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 मॉडेल्स सर्व कंपन्या लोकांना फुकट का वापरायला देतील? त्यांच्या प्रशिक्षणासाठी खर्च केलेले करोडो डॉलर्स मग त्या परत कसे मिळवणार? ते धंद्याला मारकच नाही का? तर यावर उपाय काय, हा कळीचा मुद्दा आहे.

तिसरे तत्व म्हणजे 'डेटा'. जेन-आय मॉडेल्सच्या प्रशिक्षणासाठी वापरण्यात आलेला डेटा कोठून आणला आहे? त्याच्या मालकांची परवानगी घेतली आहे का? त्यात काही पक्षपाती (बायस) तत्वे दडली आहेत का? त्यात कोणाची खाजगी अथवा गोपनीय माहिती आहे का? असे असंख्य प्रश्न निर्माण होतात. 'जसा डेटा तसे मॉडेल' हे ब्रीद असल्याने डेटा-तपासणी (ऑडिट) अत्यंत महत्वाचा मुद्दा नियमकांना विचारात घ्यावा लागणार आहे.

चौथे तत्व 'अंभालबजावणी' विषयी. मार्गदर्शक तत्वे बनवली तरी ती कशी वापरणार? प्रत्येक नवीन मॉडेलची तपासणी करून ते योग्य-सुरक्षित असल्याचे प्रमाणित (सर्टिफिकेट) करणार का? डेटा ऑडिटिंग कसे करणार? त्याच्यासाठी लागणाऱ्या संगणकीय प्रणाली (टूल्स) कोण बनवणार? आणि हे सर्व करण्यासाठी एआय-प्रशिक्षित मनुष्यबळ कसे निर्माण करणार? असे विविध प्रश्न आहेत. मॉडेल्स ओपन-सोर्स असतील तर या संदर्भात जनसामान्यांची मोठी मदत होऊ शकते. न्यूरल नेटवर्कची आणि वापरण्यात आलेल्या डेटाची माहिती सार्वजनिक केली असल्यास त्याविषयी प्रशिक्षण देणे आणि तपासणी-टूल्स बनवणे सुलभ होऊन जाईल.

अशा सर्व मुद्द्यांचा सखोल विचार करून, जागतिक सहकार्याने, विविध क्षेत्रातील तज्ज्ञांची मते जाणून घेऊन नियामक आराखडा बनवावा लागणार आहे. एकुणातच विचार करीता असे वाटते की, एआय वर नियंत्रण तर नक्कीच असावे पण त्यासाठी केलेली नियमांची चौकट व आखलेली लक्ष्मणरेखा ही काळानुसार बदलणारी, नूतन तंत्रज्ञानानुसार लवचिक होणारी आणि सुधारणेस वाव असणारी असावी, ही अपेक्षा.

- डॉ. योगेश कुलकर्णी

yogeshkulkarni@yahoo.com

