SABEU QUÈ ÉS L'ENVERINAMENT DE DADES O “DATA POISONING”?

- **Aquest implica la contaminació deliberada de les dades d’entrenament d’un sistema d’IA, introduint informació corrupta per comprometre el rendiment i resultats de la IA generativa.**

- **La principal raó per la qual s’utilitza aquesta tècnica actualment és perquè molts sistemes d’IA han estat entrenats agafant de forma indiscriminada dades en línia (especialment imatges), moltes de les quals poden estar sota copyright.**

- **Això ha donat lloc a una gran quantitat de casos d'infracció del copyright en què, els artistes en particular, han acusat a les grans empreses tecnològiques de robar i treure profit del seu treball. D’aquí que s’hagin creat eines com Nightshade per lluitar-hi de manera activa.**

- **Aquestes eines funcionen principalment alterant subtilment els píxels d'una imatge de manera que causa problemes en el retorn, si bé deixa la imatge inalterada als ulls d'una persona. Així doncs, si una organització utilitza aquestes imatges per entrenar un futur model el seu conjunt de dades pateix un enverinament o “data poisoning”. Això és perquè l'alteració de les dades originals pot provocar que l'algoritme aprengui erròniament a classificar una imatge com una cosa que no és i que una persona sabria fàcilment que no és certa.**

- **A través de l'enverinament, el sistema d’IA generativa comença a retornar resultats impredictibles i no desitjats i, és clar, com més gran sigui el nombre d'imatges enverinades a les dades d'entrenament, més gran serà la corrupció dels resultats.**

- **En definitiva, aquest enfocament posa de manifest un problema generalitzat d'abús de les dades i desafia la creença comuna que les dades que es troben en línia es poden utilitzar per a qualsevol propòsit que es cregui convenient en l’entrenament de sistemes d’IA generativa.**