QUÈ ÉS LA "MANIPULACIÓ PER SISTEMES D'IA"?

- Aquesta adopta moltes formes com estratègies addictives personalitzades per al consum de serveis, a aprofitar l'estat emocionalment vulnerable de persones en línia o l'explotació de biaixos.

- Una característica comuna subjacent d'aquestes estratègies és que influeixen en el comportament de les persones d'una manera no tansparent, reduint molt sovint el valor (econòmic) que aquestes poden obtenir mentre augmenten la rendibilitat de les empreses que les utilitzen.

- La manipulació es considera èticament incorrecta perquè (a) influeix en l'autonomia, llibertat o dignitat de les persones, (b) promou el benefici personal del manipulador a costa del manipulat, i (c) pot provocar un dany directe o indirecte per a la persona manipulada.

- La manipulació s'ha estudiat àmpliament a través de tota la tradició filosòfica, des Plató a Foucault passant per Rousseau, Kant i Arendt, amb el consens de que es tracta d'un comportament moralment contaminat irrespectivament de qui manipula (un individu, un govern, etc.).

- La manipulació per sistemes d'IA reprèn un antic debat sobre l'ús de la publicitat per manipular el comportament dels consumidors o bé l'ús de la propaganda amb finalitats polítiques i ideològiques molt prevalents en èpoques de conflicte o guerra per manipular l'opinió pública.

- A finals dels anys 1950, l'economista John Kenneth Galbraith ja va definir la publicitat com "la manipulació del desig del consumidor", i va argumentar que era molt sovint enganyosa, explotadora, i que vulnerava els principis d'autonomia, veracitat i respecte a les persones.

- Aquestes preocupacions van donar lloc al desenvolupament de les primeres directrius ètiques i normatives per a la publicitat, com el "Better Business Bureau’s Advertising Code of Ethics" i el "Federal Trade Commission's Truth in Advertising guidelines" als USA i arreu del món.

- Actualment, les noves esmenes de l'anomenada AI Act inclouen disposicions que cobreixen la manipulació per sistemes d'IA. Malgrat que aquest és un pas cabdal per protegir els ciutadans de la UE dels perills de la manipulació, les disposicions encara són massa vagues. Per què?

- Doncs pequè en la seva versió actual, la AI Act prohibeix utilitzar les "vulnerabilitats d'individus i grups específics de persones a causa dels seus trets de personalitat coneguts o previstos" però el text es refereix a "trets de personalitat" diverses vegades sense definició.

- A més, els trets de personalitat són només una petita part dels trets psicològics mesurables que poden ser explotats per un sistema d'IA. D'aquí que organitzacions com la OECD aboguin per la protecció de tot el perfil psicològic de la manipulació per part de sistemes d'IA.

- Mentrestant, sabem que la manca de transparència ajuda a l'èxit d'aquestes estratègies de manipulació per sistemes d'IA. Com? En molts casos, els usuaris no coneixen ni els objectius ni com s'utilitza la seva informació personal sensible per aconseguir aquests objectius.

- Així doncs, la possibilitat d'aquesta manipulació del comportament a través de sistemes d'IA requereix de polítiques que garanteixin l'autonomia i l'autodeterminació en qualsevol interacció entre persones i sistemes d'IA (aquesta no hauria de subordinar, enganyar o manipular).

- De fet, hauria de fer tot el contrari: complementar i augmentar les nostres habilitats com apunten les "Ethics guidelines for trustworthy AI" de la CE. El primer pas important per aconseguir aquest objectiu és millorar la transparència sobre l'abast i les capacitats de l'IA.