

IASEAI fa una crida a l'acció per a legisladors, acadèmics i la societat abans de la Cimera d'IA a París

L'Associació Internacional per a una Intel·ligència Artificial Segura i Ètica (IASEAI) ha conclòs avui la seva conferència inaugural amb una crida a l'acció amb deu mesures crucials.

7 de febrer, PARIS: La conferència **IASEAI '25** ha estat la reunió global més gran fins ara centrada en els darrers avenços en seguretat i ètica de la intel·ligència artificial. Celebrada a la seu de l'OCDE just abans de la Cimera d'Acció sobre IA a París, la conferència ha reunit destacats representants del món acadèmic, la societat civil, la indústria, els mitjans de comunicació i els governs.

El professor **Stuart Russell**, president de l'IASEAI i distingit professor d'Informàtica a la Universitat de Califòrnia, Berkeley, ha declarat:

"El desenvolupament d'una IA altament capaç probablement serà l'esdeveniment més important en la història de la humanitat. El món ha d'actuar decididament per assegurar que no sigui l'últim esdeveniment en la història de la humanitat. Aquesta conferència, i l'esperit de cooperació de la sèrie de cimeres sobre IA, em donen esperança; però hem de transformar l'esperança en acció, aviat, si volem que hi hagi un futur en el qual desitgem que visquin els nostres fills."

Les següents accions, derivades de la investigació i els debats mantinguts durant la conferència, ajudaran a avançar en la missió de l'IASEAI d'assegurar que els sistemes d'IA operin de manera segura i ètica.

IASEAI demana:

1. **Reconeixement de la importància dels nous desenvolupaments en IA.** Els responsables polítics han d'actuar amb una urgència proporcional al potencial transformador de la IA, a la velocitat del seu avenç i als riscos creixents que comporta per a la humanitat, especialment a mesura que les seves capacitats comencin a superar les humanes.
2. **Evitar la disrupció institucional i social provocada per la IA.** L'avenç de la IA amenaça amb desestabilitzar l'ocupació, les estructures socials, agreujar la desigualtat i comprometre greument l'ecosistema informatiu. Els responsables polítics han de prendre mesures proactives per protegir les institucions, les persones i les formes de vida, garantint que la IA enforteixi la societat en lloc de debilitar-la. Atès que aquests problemes tenen formes similars a diferents països, cal cercar solucions col·lectives.
3. **Abordar la cursa cap a la IA General (AGI).** Els incentius financers i estratègics per assolir una intel·ligència artificial general (AGI) o "superintel·ligència" porten empreses i països a reduir els estàndards de seguretat per obtenir un avantatge tecnològic i polític. Els responsables polítics haurien de prendre accions coordinades en fòrums internacionals, com l'ONU i l'OCDE, per garantir que la innovació i la competència es desenvolupin dins d'un marc de seguretat rigorós.

4. **Aglutinar els esforços de les comunitats de recerca al voltant de l'objectiu d'una IA segura i ètica.** Les amenaces creixents per a la prosperitat humana que plantegen els sistemes d'IA requereixen que investigadors en IA, ètica, ciències socials i política col·laborin i uneixin els seus esforços. Les diverses perspectives que ofereixen aquestes comunitats no estan en tensió; al contrari, totes contribueixen de manera important a garantir que els sistemes d'IA no perjudiquin la societat humana.
5. **Adopció de requisits de seguretat i ètica obligatoris.** Tot i que són lloables, els compromisos voluntaris de les empreses s'han de fer més específics i jurídicament vinculants. Aquests compromisos vinculants podrien incloure el registre obligatori dels sistemes d'IA avançats, incloent-hi l'autoregistre automàtic de les còpies de codi obert; la instal·lació d'interruptors de desactivació operables remotament; la notificació obligatòria d'incidents; la formació en ètica professional per a enginyers; i els estàndards per a l'entrenament, el disseny, el desenvolupament, les proves, l'auditoria i la transparència dels sistemes d'IA avançats. Els desenvolupadors haurien de proporcionar un cas de seguretat científicament convincent que els seus sistemes no crearan les anomenades *línies vermelles de comportament*, que delimiten conductes inacceptables.
6. **Foment de la cooperació global.** Els desenvolupaments en IA i els riscos associats tenen un impacte transfronterer, cosa que requereix una col·laboració internacional per a la recerca i la regulació de la seguretat de la IA. A més, els beneficis de la IA han de distribuir-se equitativament. En aquest sentit, donem la benvinguda a la creació del Consell Assessor de l'ONU sobre IA i la xarxa internacional d'Instituts de Seguretat de la IA.
7. **Augment del finançament públic en recerca.** L'escala del repte requereix una major inversió pública en recerca sobre seguretat i ètica de la IA. La recerca no pot continuar sent dominada per empreses amb conflictes d'interès. Els beneficis potencialment transformadors de la IA només es poden materialitzar si els avenços en capacitats van acompanyats de mètodes per garantir que els sistemes d'IA són segurs per disseny i alineats amb els interessos humans.
8. **Suport a la Fundació per a la IA.** La futura Fundació per a la IA, prevista pels amfitrions de la Cimera de París, és un pas important per garantir que la IA beneficiï la humanitat i enforteixi les capacitats locals. Els governs han de donar suport a aquesta fundació.
9. **Suport al Conveni Marc del Consell d'Europa sobre la IA.** Aquest tractat busca garantir que les activitats en tot el cicle de vida dels sistemes d'IA siguin coherents amb els drets humans, la democràcia i l'estat de dret. Mentre que el Consell administra el tractat, aquest està obert a la signatura d'estats no membres. 37 països l'han signat, incloent els Estats Units i el Regne Unit, i s'espera que més països el signin a la Cimera de París.
10. **Promoció d'un diàleg informat.** Investigadors i mitjans de comunicació han de treballar conjuntament per informar de manera objectiva i precisa el públic i els responsables polítics sobre els desenvolupaments en IA i la seva importància.

Sobre IASEAI: L'Associació Internacional per a una IA Segura i Ètica (IASEAI, per les seves sigles en anglès) és una organització sense ànim de lucre fundada per abordar els riscos i les oportunitats associats amb els avenços ràpids en IA. La seva missió és garantir que els sistemes d'IA operin de manera segura i ètica, i també donar forma a les polítiques, promoure la recerca, i construir comprensió i comunitat al voltant d'aquest objectiu.