

Consell Assessor de l'Observatori d'Ètica en Intel·ligència Artificial de Catalunya

Primera reunió

Data: 1 de desembre de 2020

Hora de començament: 17h

Hora d'acabament: 19:25h

Lloc: Virtual (Teams)

Assistència

Tots els membres del Consell Assessor (sense absències), el Director General d'Innovació i Economia Digital de la Generalitat de Catalunya, el Rector de la Universitat de Girona, i el Director de l'Observatori d'Ètica en Intel·ligència Artificial de Catalunya.

Programa del dia

17:00 Benvinguda i presentació de la iniciativa

-Sr. Daniel Marco, Director General d'Innovació i Economia Digital.

-Sr. Quim Salvi, Rector de la Universitat de Girona.

17:15 Presentació dels membres del Consell Assessor

18:00 Presentació dels principals objectius de l'OEIAC

18:15 Taula rodona

19:00 Fi de la reunió

Desenvolupament de la reunió

La primera reunió del Consell Assessor de l'OEIAC es va desenvolupar d'acord amb el guió previst. Després de la benvinguda i la presentació de la iniciativa per part del Sr. Daniel Marco, Director General d'Innovació i Economia Digital de la Generalitat de Catalunya, i del Sr. Quim Salvi, Rector de la Universitat de Girona, cadascun dels membres del Consell Assessor van realitzar una breu introducció personal incloent-hi interessos i trajectòries acadèmiques i/o professionals en relació a la intel·ligència artificial (IA) des de diferents àmbits (tecnològic, social, econòmic, humanístic, legal, etc.). En aquesta part de la reunió va quedar clarament de manifest que l'Observatori d'Ètica en Intel·ligència Artificial de Catalunya té en cadascuna de les persones que formen el Consell Assessor un alt grau de coneixement i expertesa, amb un grup de persones divers i multidisciplinari que aporta una visió estratègica i independent fonamental per adquirir una comprensió de les necessitats, oportunitats i responsabilitats que representa la instauració, cada vegada més generalitzada, de la IA tant dins el context català com a nivell internacional.

Un cop finalitzades les intervencions de cada membre del Consell Assessor (per ordre alfabètic), el Director de l'OEIAC va fer una breu presentació sobre els principals objectius i àmbits d'actuació d'aquesta iniciativa nascuda a finals del mes de juny de 2020, i que s'insereix en l'eix sobre "Ètica i Societat" de l'Estratègia d'Intel·ligència Artificial (IA) impulsada per la Generalitat de Catalunya amb el nom de CATALONIA.IA, i coincideix amb el Pla Estratègic UdG2030: la Suma d'Intel·ligències de la Universitat de Girona.

En finalitzar aquesta presentació es va obrir un torn obert de paraules o taula rodona per tal que els membres dels Consell Assessor aportessin temes de reflexió, idees, comentaris o impressions sobre l'ètica de la IA, i també sobre els objectius i actuacions proposades entorn a l'OEIAC. A continuació es resumeixen alguns dels temes que van sortir durant la taula rodona i pluja d'idees:

-La IA ètica cal plantejar-la tenint en compte la importància del context ja que els principis ètics són susceptibles de variar segons les percepcions, actituds, discussions i acceptació de l'ús de la IA a diferents llocs. Així doncs les normes i valors en ús de cada lloc apareixen com a fonamentals per l'aplicació de principis ètics, però també de l'estat de la tecnologia, models econòmics, sentiments de la societat civil, característiques legislatives, etc.

-Existeix una necessitat ineludible de major representació de les dones i de minories en tots els àmbits de desenvolupament i implantació de la IA (principalment a la indústria i a la universitat). Aquesta és un apartat que malgrat està canviant degut a un major promoció d'iniciatives, principalment en termes d'igualtat de gènere, per fomentar les vocacions en ciències, tecnologia, enginyeries i matemàtiques i garantir les mateixes oportunitats d'èxit laboral per noies i nois, encara queda molt recorregut per aconseguir un grau de representativitat paritària. Es fa notar que aquesta situació també pot incidir en la institucionalització de biaixos cognitius (per exemple, en termes de sexisme o racisme) per part de les persones que s'ocupen de desenvolupar o utilitzar sistemes de IA.

-La narrativa de la IA condiona l'opinió i el discurs públic sobre el desenvolupament i implantació de la IA ja que en aquesta no tan sols afloren futurs alternatius i debats sobre ells, també es poden crear falses expectatives i percepcions difícils d'invertir. En aquest sentit, les narracions que no estan estretament relacionades amb la ciència o la tecnologia poden afectar les percepcions i els graus de confiança en les possibles aplicacions i en aquells que les desenvolupen, les promouen o s'hi oposen. D'aquí que les representacions artístiques de la IA (per exemple, en termes literaris, cinematogràfics, etc.) siguin una part necessària de l'observació per la manera peculiar, de vegades contemplativa, amb que qualsevol obra artística (especialment si és bona) ens "mostra" o ens fa viure i experimentar intensament les realitats que representa.

-Es considera fonamental per avançar en la IA poder establir relacions de causalitat. Aquest és un pas important per al desenvolupament i progrés genuí de la IA ja que fins a l'actualitat molts mètodes estan focalitzats en tasques relacionades amb establir correlacions o associacions però no pas relacions de causa-efecte. En aquest sentit es preveu que l'aplicació de la inferència causal no només crearà models i prediccions més robustes, sinó que també permetrà un raonament causal i una explicabilitat de nivell superior.

-Les mètriques per mesurar l'aplicació de diferents principis ètics són necessàries però no deixen de ser aproximacions a preocupacions segons el context. També cal tenir en compte que tendeixen a posar un èmfasi excessiu en les preocupacions a curt termini. En aquest sentit cal ser flexibles i utilitzar les més representatives, també evitant una sobrerrepresentació de principis i considerant que les mètriques per avaluar l'ètica de la IA formen part d'un procés iteratiu. En qualsevol cas aquestes han de permetre, juntament amb avaluacions qualitatives, l'obtenció d'una imatge més completa de l'estat de l'ètica de la IA a nivell general, i particularment detectar les implicacions per a les persones

interessades, principalment les més afectades o vulnerables per l'ús o abús d'un sistema de IA.

-Cal una millor capacitat de les persones per governar la IA, i cal que els processos que governen la IA siguin observables, però no tan sols observar, cal actuar i promoure bones pràctiques establint prioritats i exemples concrets de transparència. En aquest sentit un aspecte que apareix com a important és el de la fiabilitat o, dit d'altra manera, si el sistema de IA realitza la tasca que se li encomana d'una manera fiable i suficientment transparent per tal de verificar que el sistema fa el que preteníem. De vegades, però, els processos són tan complexos que simplement no és possible trobar una explicació de per què el sistema decideix una i no una altra cosa. Així doncs es subratlla que cal pensar maneres d'explicar diferents parts o fases dels processos utilitzats ja que d'això depenen qüestions ètiques com la responsabilitat, la rendició de comptes i també de qualitat i significació de la mateixa presa de decisions.

-En el camp de la governança de la IA es subratlla que hi ha una creixent necessitat d'incloure mecanismes de control i auditoria per part de l'administració com, per exemple, representen els registres d'algoritmes engegats per part d'algunes ciutats europees com Helsinki i Amsterdam. Aquestes ciutats han instaurat registres d'IA per detallar com cada govern de la ciutat utilitza algoritmes per proporcionar serveis als seus ciutadans. La idea és clarament de governança de la IA a nivell local però traslladable a nivells regionals, estatals i supranacionals. L'objectiu és proporcionar transparència a través de descripcions detallades de com s'utilitza un determinat algoritme, així com els possibles biaixos o riscos. Un element que es comenta és la possibilitat que aquests registres permetin empoderar els ciutadans i donar-los una manera d'avaluar, examinar o qüestionar les aplicacions d'IA. No queda clar, però, com es pot aplicar un model semblant de registre d'algoritmes per l'àmbit privat. Potser a través dels registres de patents, malgrat molts algoritmes no cal que estiguin patentats per ser utilitzats.

-La naturalesa de les dades que s'utilitzen en els sistemes de IA es considera una part important del problema actual, principalment en termes de privacitat i seguretat, però també en relació a la confiança, ja que moltes dades (principalment massives) s'utilitzen sense consentiment de les persones que les generen en primera instància. Un altre tema relacionat amb les dades és que aquestes estan subjectes en molts casos als biaixos de recollida, fet que posa en dubte la pròpia legitimitat del procés de recollida i d'utilització. En aquest sentit, cal una major representativitat en la recollida d'informació per tal que els sistemes de IA no reproduïxin solucions amb biaix, i una mitigació constant per tal que biaixos ineludibles puguin ser tractats a través de la gestió de les dades.

-No totes les empreses que utilitzen IA tenen com a finalitat maximitzar beneficis, el desenvolupament sostenible també està present en la seva agenda i la prova és que fan cada cop més referència a la competitivitat de la IA ètica (sovint des d'una perspectiva molt pragmàtica). Es subratlla que aquest és un aspecte positiu però en alguns o molts casos, incloses en les anomenades "Big Tech", el fet que les empreses tinguin iniciatives de IA ètica no significa que vagin més enllà d'una adopció comercial, o sigui sense aplicació. En aquest sentit s'apunta el que es coneix com a "ethics washing" en l'àmbit de la IA com una estratègia errònia o de manca de voluntat real per afrontar els diferents reptes ètics que planteja el desenvolupament i implantació de la IA.

-Finalment, s'esmenta el paper de la administració pública en l'adopció de sistemes d'IA per a la provisió de serveis a la ciutadania. En aquest sentit es fa referència a l'explotació de les dades massives administratives que, en molts casos, no estan recollides, ni tractades ni analitzades amb sistemes d'IA. Malgrat això, és evident que existeix un interès creixent tant per part de l'administració pública com per l'empresa privada d'explotar i maximitzar aquests recursos de dades massives de caire administratiu. Sens dubte, aquest escenari representa diferents reptes al voltant de la IA ètica, principalment en termes de responsabilitat, justícia i seguretat o privacitat.