



Il-Grupp ta' Esperti ta' Livell Għoli

DWAR L-INTELLIĠENZA ARTIFICJALI TAL-KUMMISSJONI



ABBOZZ TAL-LINJI GWIDA DWAR L-ĒTIKA GĦAL IA AFFIDABBLI SOMMARJU EŻEKUTTIV

Dokument ta' Hidma għall-konsultazzjoni tal-partijiet
ikkoncernati

ABBOZZ TAL-LINJI GWIDA DWAR L-ĒTIKA

GĦAL IA AFFIDABBLI



Grupp ta' Esperti ta' Livell Għoli dwar l-Intelliġenza Artifiċjali
Abbozz tal-Linji Gwida dwar l-Etika għal IA Affidabbli

Il-Kummissjoni Ewropea
Id-Direttorat Ġenerali għall-Komunikazzjoni

Kuntatt Nathalie Smuha - Koordinatur tal-AI HLEG
Indirizz elettroniku CNECT-HLG-AI@ec.europa.eu

Il-Kummissjoni Ewropea
B-1049 Brussell

Id-dokument sar pubbliku fit-18 ta' Diċembru 2018, bl-Ingliż.

Dan id-dokument ta' ħidma ġie prodott mill-AI HLEG mingħajr preġudizzju għall-pożizzjoni individwali tal-membri tiegħu dwar punti speċifiċi, u mingħajr preġudizzju għall-verżjoni finali tad-dokument. Dan id-dokument se jibqa' jiġi żviluppat u verżjoni finali tiegħu se tiġi ppreżentata f'Marzu 2019 wara konsultazzjoni mal-partijiet interessati permezz tal-Alleanza Ewropea għall-IA.

La l-Kummissjoni Ewropea u l-ebda persuna li tagħxi f'isem il-Kummissjoni ma hija responsabbli għall-użu li jista' jsir mit-tagħrif li ġej. Il-kontenut ta' dan id-dokument ta' ħidma huwa r-responsabbiltà unika tal-Grupp ta' Esperti ta' Livell Għoli dwar l-Intelliġenza Artifiċjali (AI HLEG). Għalkemm il-persunal tas-servizzi tal-Kummissjoni ffaċilita t-tnejn tal-Linji Gwida, l-opinjoni espressi f'dan id-dokument jirriflettu l-opinjoni tal-AI HLEG, u ma jstgħu fl-ebda ċirkostanza jitqiesu li qed jiddikjaraw pożizzjoni uffiċjali tal-Kummissjoni Ewropea. Dan huwa abbozz tal-ewwel Riżultat tal-AI HLEG. Verżjoni finali tiegħu se tiġi ppreżentata lill-Kummissjoni f'Marzu 2019. Verżjoni finali tat-tieni Riżultat – ir-Rakkomandazzjonijiet ta' Politika u Investiment tal-IA – se jiġu ppreżentati f'nofs l-2019.

Aktar informazzjoni dwar il-Grupp ta' Esperti ta' Livell Għoli dwar l-Intelliġenza Artifiċjali hija disponibbli online (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>). Il-politika tal-użu mill-ġdid tad-dokumenti tal-Kummissjoni Ewropea hija rregolata bid-Deċiżjoni 2011/833/UE (ĠU L 330, 14.12.2011, p.39). Għal kull użu jew riproduzzjoni ta' ritratti jew ta'

materjal ieħor li mhuwiex kopert mid-drittijiet tal-awtur tal-UE, irid jitalab permess direttament mingħand id-detenturi tad-drittijiet tal-awtur.

SOMMARJU EŻEKUTTIV

Dan id-dokument ta' ħidma jikkostitwixxi abbozz tal-Linji Gwida dwar l-Etika fl-IA li ġew imfassla mill-Grupp ta' Ħidma ta' Livell Għoli tal-Kummissjoni Ewropea dwar l-Intelliġenza Artifiċjali (AI HLEG), li l-verżjoni finali tagħhom hija dovuta f'Marzu 2019.

L-Intelliġenza Artifiċjali (IA) hija waħda mill-aktar forzi trasformattivi b'saħħitha ta' żmienna, u hija ddestinata li tibdel in-nisġa tas-soċjetà. Hija tipprezenta opportunità kbira biex jiżdiedu l-prosperità u t-tkabbir, li l-Ewropa trid tistinka biex tikseb. F'dan l-aħħar deċennju, twettqu avvanzi ewlenin minħabba d-disponibbiltà ta' ammonti kbar ta' *data* digitali, arkitetturi tal-computing potenti, u avvanzi fit-tekniki tal-IA bħat-tagħlim awtomatiku. L-iżviluppi ewlenin li twettqu permezz tal-IA fil-vetturi awtonomi, fil-kura tas-saħħa, fir-robots tad-dar/ta' servizz, fl-edukazzjoni jew fiċ-ċibersigurtà qed itejbu l-kwalità ta' ħajjitna kuljum. Barra minn hekk, l-IA hija kruċjali biex jiġu indirizzati ħafna mill-isfidi l-kbar li qed tiffaċċja d-dinja, bħas-saħħa u l-benessri globali, it-tibdil fil-klima, sistemi legali u demokratiċi affidabbli u sfidi oħrajn li huma espressi fl-Għanijiet ta' Żvilupp Sostenibbli tan-Nazzjonijiet Uniti.

Filwaqt li għandha l-kapaċità li tiġġenera benefiċċji kbar għall-individwi u għas-soċjetà, l-IA tohloq ukoll ċerti riskji li għandhom jiġu ġestiti b'mod xieraq. Peress li, b'mod globali, il-benefiċċji tal-IA huma akbar mir-riskji tagħha, jeħtieġ li nsegwu t-triq li **timmassimizza l-benefiċċji tal-IA filwaqt li timminimizza r-riskji tagħha**. Sabiex niżguraw li nibqgħu fit-triq it-tajba, **jenħtieġ approċċ għall-IA li jkun iċċentrat fuq il-bniedem**, li jobligana nżommu f'moħħna li l-iżvilupp u l-użu tal-IA m'għandhomx jittqiesu mezz fihom infushom, iżda li għandhom l-għan li jtejjbu l-benessri tal-bniedem. **L-IA affidabbli se tkun il-kewkba polari tagħna**, għaliex il-bniedem jistgħu biss jiksbu b'kunfidenza u b'mod sħiħ il-benefiċċji tal-IA jekk ikunu jistgħu jafdaw it-teknoloġija.

L-IA affidabbli għandha **żewġ komponenti**: (1) jenħtieġ li tirrispetta d-drittijiet fundamentali, ir-regolamenti applikabbli u l-prinċipji u l-valuri ewlenin, li jiżguraw "**għan etiku**" u (2) jenħtieġ li tkun **teknikament b'saħħitha** u affidabbli peress li, anke b'intenzjonijiet tajbin, in-nuqqas ta' għarfien teknoloġiku jista' jikkawża ħsara mhux intenzjonata.

Dawn il-Linji Gwida għalhekk jistabbilixxu **qafas għal IA Affidabbli**:

- **Il-Kapitolu I** jitratta **l-iżgurar tal-għan etiku tal-IA**, billi jistabbilixxi d-drittijiet fundamentali, il-prinċipji u l-valuri li għandu jikkonforma magħhom.
- Minn dawk il-prinċipji, **il-Kapitolu II** joħroġ **gwida dwar it-twettiq** ta' IA Affidabbli, li tittratta kemm l-għan etiku kif ukoll is-saħħa teknika. Dan isir bl-elenkar tar-rekwiżiti għal IA Affidabbli u li joffri ħarsa ġenerali lejn il-metodi tekniċi u mhux tekniċi li jistgħu jintużaw għall-implimentazzjoni tiegħu.
- **Il-Kapitolu III** sussegwentement **joperazzjonalizza** r-rekwiżiti billi jipprovdi lista ta' valutazzjoni konkreta iżda mhux eżawrjenti għal IA Affidabbli. Din il-lista mbagħad tiġi adattata għal każijiet ta' użu speċifiku.

B'kuntrast ma' dokumenti oħra li jittrattaw l-IA etika, il-Linji Gwida għalhekk ma għandhomx l-għan li jipprovdu lista oħra ta' valuri u prinċipji ewlenin għall-IA, iżda pjuttost li joffru gwida għall-implimentazzjoni u l-operazzjonalizzazzjoni konkreta tagħhom f'sistemi ta' IA. Din il-gwida tiġi pprovduta bi tliet saffi ta' astrazzjoni, mill-aktar astratt fil-Kapitolu I (drittijiet, prinċipji u valuri fundamentali), sal-aktar konkret fil-Kapitolu III (lista ta' valutazzjoni).

Il-Linji Gwida huma indirizzati lill-**partijiet ikkonċernati rilevanti kollha li qed jiżviluppaw, jimplimentaw jew jużaw l-IA**, li jinkludu kumpaniji, organizzazzjonijiet, riċerkaturi, servizzi pubbliċi, istituzzjonijiet, individwi jew entitajiet oħra. Fil-verżjoni finali ta' dawn il-Linji Gwida, se jiġi pprezentat mekkanizmu li jippermetti lill-partijiet ikkonċernati japprovawhom b'mod volontarju.

Importanti jingħad li dawn il-Linji Gwida mhumiex intiżi bħala sostitut għal kwalunkwe forma ta' tfassil ta' politika jew regolamentazzjoni (li għandhom jiġu indirizzati fit-tieni riżultat tal-AI HLEG: ir-Rakkomandazzjonijiet ta' Politika u Investment, mistennija f'Mejju 2019), u lanqas ma għandhom l-għan li jiskoraġġixxu l-introduzzjoni tagħhom. Barra minn hekk, il-Linji Gwida għandhom jitqiesu bħala dokument ħaj li jeħtieġ li jiġi aġġornat regolarment maż-żmien sabiex tiġi żgurata rilevanza kontinwa hekk kif tevolvi t-teknoloġija u l-għarfien tagħna fir-rigward tagħha. Dan id-dokument għandu għalhekk ikun punt tat-tluq għad-diskussjoni dwar **"IA Affidabbli magħmula fl-Ewropa"**.

Għalkemm l-Ewropa tista' biss ixxandar l-approċċ etiku tagħha lejn l-IA meta dan ikun kompetittiv fil-livell globali, **approċċ etiku għall-IA huwa kruċjali biex jippermetti kompetittività responsabbli**, għaliex dan jiġġenera l-fiducia tal-utenti u jiffacilita l-adozzjoni aktar wiesgħa tal-IA. Dawn il-Linji Gwida mhumiex intenzjonati li joħonqu l-innovazzjoni tal-IA fl-Ewropa, iżda minflok għandhom l-għan li jużaw l-etika bħala ispirazzjoni biex jiżviluppaw marka unika tal-IA, waħda li jkollha l-għan li tiproteġi u li tkun ta' benefiċċju kemm għall-individwi kif ukoll għall-ġid komuni. Dan jippermetti lill-Ewropa tippożizzjona lilha nnifisha bħala mexxej f'IA avanzata, sigura u etika. Iċ-ċittadini Ewropej jistgħu biss igawdu mill-benefiċċji tal-IA b'mod sħiħ billi tiġi żgurata l-fiducia fiha.

Fl-aħħar nett, lil hinn mill-Ewropa, dawn il-Linji Gwida għandhom l-għan ukoll li **jrawmu riflessjoni u diskussjoni** dwar qafas etiku għall-IA fil-livell globali.

GWIDA EŻEKUTTIVA

Kull Kapitolu tal-Linji Gwida joffri gwida dwar kif naslu għal IA affidabbli, gwida indirizzata lill-partijiet ikkonċernati kollha rilevanti li jiżviluppaw, jimplimentaw jew jużaw l-IA, imqassra hawn taħt:

Kapitolu I: Gwida Ewlenija għall-Iżgurar ta' Għan Etiku:

- Żgura li l-IA tkun **iċċentrata fuq il-bniedem**: L-IA għandha tiġi żviluppata, implimentata u użata b'"**għan etiku**", li jirrifletti u jissejjes fuq id-drittijiet fundamentali, il-valuri tas-soċjetà u l-prinċipji etici tal-*Benefiċjenza* (li tagħmel tajjeb), in-*Nonkrudeltà* (li ma tagħmilx ħsara), *l-Awtonomija tal-bnedmin*, *il-Ġustizzja*, u *l-Ispjegabbiltà*. Dan huwa ta' importanza kruċjali għall-ħidma lejn **IA Affidabbli**.
- Ibbaza ruħek fuq id-drittijiet fundamentali, il-prinċipji etici u l-valuri biex prospettivament tevalwa l-effetti possibbli li l-IA jista' jkollha fuq il-bnedmin u l-ġid komuni. Agħti **attenzjoni partikolari** lis-sitwazzjonijiet li jinvolvu aktar **gruppi vulnerabbli** bħat-tfal, il-persuni b'diżabbiltà jew il-minoranzi, jew lis-sitwazzjonijiet b'**asimetriji ta' setgħa jew informazzjoni**, bħal pereżempju dawk bejn min iħaddem u l-impjegati, jew bejn in-negozji u l-konsumaturi.
- Agħraf u kun konxju tal-fatt li, filwaqt li ġġib benefiċċji sostanzjali lill-individwi u s-soċjetà, l-IA jista' jkollha wkoll impatt negattiv. Ibqa' vigili għal oqsma ta' tħassib kbir.

Kapitolu II: Gwida Ewlenija għar-Realizzazzjoni ta' IA Affidabbli:

- Inkorpora **r-rekwiżiti għal IA Affidabbli mill-aktar fażi bikrija tad-disinn**: Responsabbiltà, Governanza tad-*Data*, Disinn għal kulħadd, Governanza tal-IA Awtonoma (Superviżjoni tal-Bniedem), Non-Diskriminazzjoni, Rispett tal-Awtonomija Umana, Rispett tal-Privatezza, Robustezza, Sikurezza, Trasparenza.
- Ikkunsidra metodi tekniċi u mhux tekniċi biex tiżgura l-implimentazzjoni ta' dawk ir-rekwiżiti fis-sistema tal-IA. Barra minn hekk, żomm dawk ir-rekwiżiti f'moħħok meta tkun qed tgħaqqad tim biex jaħdem fuq is-sistema, is-sistema fiha nnifisha, l-ambjent ta' ttestjar u l-applikazzjonijiet potenzjali tas-sistema.

- Ipprovi, b' mod ċar u proattiv, informazzjoni lill-partijiet ikkonċernati (konsumaturi, impjegati, eċċ.) dwar il-kapaċitajiet u l-limitazzjonijiet tas-sistema ta' IA, li tippermettilhom jstabilixxu aspettattivi realistiċi. L-iżgurar tat-**Traċċabbiltà** tas-sistema tal-IA huwa essenzjali f'dan ir-rigward.
- Agħmel l-IA Affidabbli **parti mill-kultura tal-organizzazzjoni**, u pprovi informazzjoni lill-partijiet ikkonċernati dwar kif l-IA Affidabbli tiġi implimentata fid-disinn u l-użu tas-sistemi ta' IA. L-IA Affidabbli tista' tiġi inkluża wkoll fil-karti ta' deontoloġija jew fil-kodicijiet ta' kondotta tal-organizzazzjonijiet.
- Żgura l-partecipazzjoni u l-**inklużjoni tal-partijiet ikkonċernati** fid-disinn u fl-iżvilupp tas-sistema ta' IA. Barra minn hekk, żgura d-**diversità** meta twaqqaf it-timijiet li jiżviluppaw, jimplementaw u jittestjaw il-prodott.
- Stinka biex **tiffaċilita l-possibbiltà ta' awditjar** tas-sistemi tal-IA, b' mod partikolari f'kuntesti jew f'sitwazzjonijiet kritiċi. Kemm jista' jkun, iddisinja s-sistema tiegħek b'tali mod li tippermetti l-intraċċar ta' deċiżjonijiet individwali għal diversi inputs tiegħek; mudelli tad-*data*, mudelli mħarrġin minn qabel, eċċ. Barra minn hekk, iddefinixxi l-**metodi ta' spjegazzjoni** tas-sistema tal-IA.
- Żgura proċess speċifiku għall-**governanza tar-responsabbiltà**.
- Aħsiblu minn qabel għat-**taħriġ u l-edukazzjoni**, u żgura li l-manigers, l-iżviluppaturi, l-utenti u l-impjegaturi jkunu konxji tal-IA Affidabbli u mħarrġa fiha.
- Kun konxju li jista' jkun hemm tensjonijiet fundamentali bejn l-oġettivi differenti (it-trasparenza tista' tiftaħ il-bieb għal użu ħażin; l-identifikazzjoni u l-korrezzjoni tal-preġudizzju jistgħu jikkuntrastaw mal-protezzjonijiet tal-privatezza). Ikkomunika u ddokumenta dawn il-kompromessi.
- Rawwem ir-riċerka u l-innovazzjoni biex tippromwovi l-kisba tar-rekwiziti għal IA Affidabbli.

Kapitolu III: Gwida Ewlenija għall-Valutazzjoni ta' IA Affidabbli

- Adotta **lista ta' valutazzjoni** għal IA Affidabbli meta tiżviluppa, timplementa, jew tuża l-IA, u adattatha għall-każ speċifiku li fih tkun qed tintuża s-sistema.
- Żomm f'moħħok li lista ta' valutazzjoni qatt **mhi se tkun eżawrjenti**, u li l-iżgurar ta' IA Affidabbli m'għandhiex x'taqsam mal-immarrar tal-kaxxi, iżda ma' proċess kontinwu tal-identifikazzjoni tar-rekwiziti, l-evalwazzjoni tas-soluzzjonijiet u l-iżgurar ta' eżiti mtejba tul iċ-ċiklu kollu tal-ħajja tas-sistema tal-IA.

Din il-gwida tiffurma parti minn viżjoni li tħaddan approċċ iċċentrat fuq il-bniedem għall-Intelliġenza Artifiċjali, li se jippermetti lill-Ewropa tkun innovatur globali ewleni fl-IA etika, sigura u avvanzata. Hija tħabrek biex tiffaċilita u tippermetti **"IA Affidabbli magħmula fl-Ewropa"** li se ssaħħaħ il-benesseri taċ-ċittadini Ewropej.