



Bruxelles, 8.4.2019
COM(2019) 168 final

**COMUNICAZIONE DELLA COMMISSIONE AL PARLAMENTO EUROPEO,
AL CONSIGLIO, AL COMITATO ECONOMICO E SOCIALE EUROPEO E
AL COMITATO DELLE REGIONI**

Creare fiducia nell'intelligenza artificiale antropocentrica

1. INTRODUZIONE – LA STRATEGIA EUROPEA PER L'INTELLIGENZA ARTIFICIALE

L'intelligenza artificiale (IA) può trasformare il nostro mondo in positivo: può migliorare l'assistenza sanitaria, ridurre il consumo di energia, rendere le automobili più sicure e consentire agli agricoltori di utilizzare le risorse idriche e naturali in modo più efficiente. L'IA può essere utilizzata per prevedere i cambiamenti ambientali e climatici, migliorare la gestione dei rischi finanziari e fornire gli strumenti per fabbricare con minor spreco prodotti su misura per le nostre esigenze; può anche contribuire a intercettare le frodi e le minacce di cibersicurezza e consente alle forze dell'ordine di lottare più efficacemente contro la criminalità.

L'IA può portare benefici all'intera società e all'economia. È una tecnologia strategica attualmente in fase di sviluppo e di diffusione a ritmi rapidi in tutto il mondo. Crea però anche nuove sfide per il futuro del lavoro e solleva questioni giuridiche ed etiche.

Per affrontare queste sfide e sfruttare al massimo le opportunità offerte dall'IA, nell'aprile 2018 la Commissione ha pubblicato una strategia europea¹ che pone l'essere umano al centro dello sviluppo dell'IA – un'IA antropocentrica. Si tratta di un approccio su tre fronti volto a dare impulso alla capacità tecnologica e industriale dell'UE e all'adozione dell'IA in tutti i settori economici, a prepararsi ai cambiamenti socio-economici e ad assicurare un quadro etico e giuridico adeguato.

Per raggiungere gli obiettivi della strategia per l'IA, **la Commissione ha sviluppato di concerto con gli Stati membri un piano coordinato sull'IA²** che ha presentato nel dicembre 2018 con lo scopo di creare sinergie, condividere i dati – la materia prima per molte applicazioni di IA – e aumentare gli investimenti congiunti. L'obiettivo è promuovere la cooperazione transfrontaliera e mobilitare tutti gli attori per aumentare gli investimenti pubblici e privati ad **almeno 20 miliardi di EUR** l'anno nei prossimi dieci anni.³ La Commissione ha raddoppiato gli investimenti nell'IA nel quadro di Orizzonte 2020 e prevede di investire un miliardo di EUR l'anno a titolo dei programmi Orizzonte Europa e Europa digitale, soprattutto per sostenere spazi di dati comuni nei settori della sanità, dei trasporti e delle attività produttive, grandi centri di sperimentazione come ospedali intelligenti e infrastrutture per i veicoli automatizzati e un'agenda strategica di ricerca.

Per attuare tale agenda strategica comune di ricerca, innovazione e diffusione dell'IA, la Commissione ha intensificato il **dialogo con tutti i pertinenti portatori di interessi** dell'industria, degli istituti di ricerca e delle autorità pubbliche. Il nuovo programma Europa digitale sarà inoltre fondamentale per aiutare a mettere l'IA a disposizione delle piccole e medie imprese in tutti gli Stati membri attraverso poli di innovazione digitale, strutture di prova e sperimentazione potenziate, spazi di dati e programmi di formazione.

Forte della sua reputazione per prodotti sicuri e di alta qualità, l'Europa adotta un approccio etico all'IA che rafforza la fiducia dei cittadini nello sviluppo digitale e mira a creare un vantaggio competitivo per le imprese europee attive nel campo dell'IA. Obiettivo della presente comunicazione è dare il via a una fase pilota esaustiva che coinvolga i portatori di interessi sulla più ampia scala nel testare l'attuazione pratica degli orientamenti etici per lo sviluppo e l'utilizzo dell'IA.

¹ COM(2018) 237.

² COM(2018) 795.

³ Per contribuire al raggiungimento di questo obiettivo, nell'ambito del prossimo periodo di programmazione 2021-2027 la Commissione ha proposto che l'Unione assegni almeno un miliardo di EUR in finanziamenti a titolo dei programmi Orizzonte Europa e Europa digitale per investire nell'IA.

2. CREARE FIDUCIA NELL'IA ANTROPOCENTRICA

La strategia europea per l'IA e il piano coordinato sull'IA indicano chiaramente che **la fiducia è una condizione indispensabile per assicurare un approccio antropocentrico all'IA**: l'intelligenza artificiale non è fine a se stessa, ma è uno strumento a servizio delle persone che ha come fine ultimo quello di migliorare il benessere degli esseri umani. Per questo occorre **garantire l'affidabilità dell'IA**. I valori su cui si basano le nostre società devono essere pienamente integrati nelle modalità di sviluppo dell'IA.

L'Unione si fonda sui **valori del rispetto della dignità umana, della libertà, della democrazia, dell'uguaglianza, dello Stato di diritto e del rispetto dei diritti umani**, compresi i diritti delle persone appartenenti a minoranze⁴. Questi valori sono comuni agli Stati membri in una società caratterizzata dal pluralismo, dalla non discriminazione, dalla tolleranza, dalla giustizia, dalla solidarietà e dalla parità tra donne e uomini. Inoltre, la **Carta dei diritti fondamentali dell'UE** raccoglie in un unico testo tutti i diritti personali, civici, politici, economici e sociali delle persone all'interno dell'UE.

L'UE dispone di un **solido quadro normativo** destinato a diventare lo standard internazionale per un'IA antropocentrica. Il regolamento generale sulla protezione dei dati garantisce un elevato livello di protezione dei dati personali e prevede l'attuazione di misure per assicurare una protezione dei dati fin dalla progettazione e per impostazione predefinita⁵. Il regolamento sulla libera circolazione dei dati non personali elimina gli ostacoli alla libera circolazione dei dati non personali e garantisce il trattamento di tutte le categorie di dati ovunque in Europa. Il regolamento sulla cibersicurezza, adottato di recente, contribuirà a rafforzare la fiducia nei confronti del mondo online e anche la proposta di regolamento sulla vita privata elettronica (e-privacy)⁶ ha lo stesso obiettivo.

L'IA pone tuttavia nuove sfide perché consente alle macchine di "imparare", prendere decisioni ed eseguirle senza l'intervento umano. In breve tempo questa modalità di funzionamento diventerà la norma per molti tipi di beni e servizi, dagli smartphone alle automobili automatizzate, ai robot e alle applicazioni online. Le decisioni prese dagli algoritmi potrebbero però basarsi su dati incompleti e quindi non affidabili, manomessi a seguito di attacchi informatici, inficiati da condizionamenti o semplicemente errati. Applicare acriticamente la tecnologia man mano che questa viene sviluppata porterebbe pertanto a risultati problematici e a una certa riluttanza da parte dei cittadini ad accettarla o a utilizzarla.

La tecnologia dell'IA dovrebbe invece essere sviluppata in modo da porre al centro l'essere umano e permetterle di conquistare la fiducia del pubblico. Di conseguenza, le applicazioni di IA dovrebbero non solo rispettare la legge, ma anche osservare i principi etici e garantire che le loro attuazioni pratiche non comportino danni indesiderati. La diversità in termini di sesso, razza o origine etnica, religione o convinzioni personali, disabilità ed età dovrebbe essere

⁴ L'Unione europea è anche parte della convenzione delle Nazioni Unite sui diritti delle persone con disabilità.

⁵ Regolamento (UE) 2016/679. Il regolamento generale sulla protezione dei dati garantisce la libera circolazione dei dati personali all'interno dell'Unione. Contiene disposizioni relative al processo decisionale basato unicamente sul trattamento automatizzato, compresa la profilazione. Gli interessati hanno il diritto di essere informati dell'esistenza di processi automatizzati nel processo decisionale e di ricevere informazioni significative sulla logica utilizzata nel processo decisionale automatizzato, come pure sull'importanza e sulle conseguenze previste del trattamento per la loro persona. In tali casi gli interessati hanno anche il diritto di ottenere l'intervento umano, di esprimere la propria opinione e di contestare la decisione.

⁶ COM(2017) 10.

garantita in ogni fase dello sviluppo dell'IA. Le applicazioni di IA dovrebbero dare potere alle persone e rispettarne diritti fondamentali; dovrebbero puntare a rafforzare le capacità dei cittadini, non a sostituirsi a loro, e consentire l'accesso anche alle persone con disabilità.

Vi è quindi la necessità di elaborare **orientamenti etici** basati sul quadro normativo esistente e che dovrebbero essere applicati da sviluppatori, fornitori e utenti dell'IA nel mercato interno, stabilendo condizioni di parità sul piano etico in tutti gli Stati membri. Per questo motivo la Commissione ha istituito un **gruppo di esperti ad alto livello sull'IA**⁷ in rappresentanza di una vasta gamma di portatori di interessi, incaricandolo della stesura di orientamenti etici sull'IA e della redazione di una serie di raccomandazioni per una più ampia politica in materia di IA. Allo stesso tempo è stata istituita l'**Alleanza europea per l'IA**⁸, piattaforma multilaterale aperta con oltre 2 700 membri, per fornire un contributo più vasto ai lavori del gruppo di esperti ad alto livello sull'IA.

Nel dicembre 2018 il gruppo di esperti ad alto livello sull'IA ha pubblicato un primo progetto di orientamenti etici; in seguito a una **consultazione dei portatori di interessi**⁹ e a **riunioni con rappresentanti degli Stati membri**¹⁰, nel marzo 2019 il gruppo di esperti sull'IA ha presentato alla Commissione un documento riveduto. Finora nel loro feedback i portatori di interessi hanno generalmente accolto con favore il carattere pratico degli orientamenti e la guida concreta che offrono agli sviluppatori, ai fornitori e agli utenti dell'IA su come garantire l'affidabilità.

2.1. Orientamenti per un'IA affidabile elaborati dal gruppo di esperti ad alto livello sull'IA

Gli orientamenti elaborati dal gruppo di esperti ad alto livello sull'IA a cui la presente comunicazione fa riferimento¹¹ si basano in particolare sul lavoro svolto dal Gruppo europeo sull'etica nelle scienze e nelle nuove tecnologie (GEE) e dall'Agenzia dell'Unione europea per i diritti fondamentali (FRA).

Gli orientamenti partono dal presupposto che per ottenere un'"intelligenza artificiale affidabile" sono necessari tre elementi: 1) l'IA dovrebbe rispettare la legge, 2) dovrebbe osservare i principi etici e 3) dovrebbe dimostrare robustezza.

Sulla base di questi tre elementi e dei valori europei illustrati nella sezione 2, gli orientamenti individuano sette requisiti fondamentali che le applicazioni di IA dovrebbero soddisfare per essere considerate affidabili. Gli orientamenti contengono anche una lista di controllo che aiuta a verificare nella pratica se tali requisiti sono soddisfatti.

I sette requisiti fondamentali sono:

- intervento e sorveglianza umani,

⁷ <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>.

⁸ <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>.

⁹ La consultazione ha permesso di raccogliere le osservazioni di 511 organizzazioni, associazioni, imprese, istituti di ricerca, singoli cittadini e altri soggetti. Una sintesi del feedback ricevuto è consultabile al seguente indirizzo: https://ec.europa.eu/futurium/en/system/files/ged/consultation_feedback_on_draft_ai_ethics_guidelines_4.pdf.

¹⁰ Il lavoro del gruppo di esperti è stato accolto positivamente dagli Stati membri. Nelle sue conclusioni del 18 febbraio 2019 il Consiglio ha preso atto tra l'altro dell'imminente pubblicazione degli orientamenti etici e sostenuto lo sforzo della Commissione teso a far emergere sulla scena mondiale l'approccio etico dell'UE: <https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/it/pdf>.

¹¹ <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>.

- robustezza tecnica e sicurezza,
- riservatezza e governance dei dati,
- trasparenza,
- diversità, non discriminazione ed equità,
- benessere sociale e ambientale,
- accountability.

Benché questi requisiti siano applicabili a tutti i sistemi di IA nei diversi contesti e settori, per la loro attuazione concreta e proporzionata è opportuno tenere conto del contesto specifico in cui si applicano e adottare un approccio basato sull'impatto. A titolo di esempio, un'applicazione di IA che suggerisce di leggere un libro non adatto comporta molti meno rischi rispetto a una che sbaglia una diagnosi di tumore e potrebbe quindi essere sottoposta a una vigilanza meno rigorosa.

Gli orientamenti elaborati dal gruppo di esperti ad alto livello sull'IA non sono vincolanti e non introducono pertanto nuovi obblighi giuridici. Tuttavia numerose disposizioni vigenti del diritto dell'Unione (spesso settoriali o specifiche per un determinato uso) già comprendono uno o più di questi requisiti fondamentali, ad esempio la sicurezza, la protezione dei dati personali, la tutela della riservatezza o le norme per la salvaguardia dell'ambiente.

La Commissione accoglie con favore il lavoro del gruppo di esperti ad alto livello sull'IA e lo considera un contributo prezioso per la propria attività di elaborazione delle politiche.

2.2. Requisiti fondamentali per un'IA affidabile

La Commissione appoggia i requisiti fondamentali per un'IA affidabile di seguito descritti, che si basano su valori europei. Incoraggia i portatori di interessi ad applicarli e a testare la lista di controllo che li inserisce in un processo operativo al fine di instaurare il giusto ambiente di fiducia per uno sviluppo e un utilizzo efficaci dell'IA. La Commissione auspica un riscontro dei portatori di interessi per valutare se la lista di controllo fornita con gli orientamenti debba essere ulteriormente adeguata.

I. Intervento e sorveglianza umani

I sistemi di IA dovrebbero aiutare le persone a compiere scelte migliori e più consapevoli nel perseguimento dei loro obiettivi. Dovrebbero promuovere lo sviluppo di una società fiorente ed equa sostenendo l'intervento umano e i **diritti fondamentali** e non ridurre, limitare o fuorviare l'autonomia umana. Il generale **benessere dell'utente** dovrebbe essere al centro della funzionalità del sistema.

La sorveglianza umana contribuisce a garantire che i sistemi di IA non mettano in pericolo l'autonomia umana o provochino altri effetti negativi. A seconda dello specifico sistema basato sull'IA e del relativo settore di applicazione si dovrebbero prevedere **misure di controllo** di livello adeguato, comprese l'adattabilità, l'accuratezza e la spiegabilità di tali

sistemi¹². La **sorveglianza** può essere effettuata mediante meccanismi di governance che garantiscano l'adozione di un approccio con intervento umano ("human-in-the-loop"), con supervisione umana ("human-on-the-loop") o con controllo umano ("human-in-command")¹³. Deve essere garantito che le autorità pubbliche siano in grado di esercitare una sorveglianza in linea con il loro mandato. A parità di altre condizioni, meno il sistema di IA è sorvegliato dall'uomo più devono essere dettagliati i test e rigorosa la governance.

II. Robustezza tecnica e sicurezza

Perché l'IA sia affidabile è indispensabile che gli algoritmi siano sicuri, affidabili e sufficientemente robusti da far fronte a errori o incongruenze durante tutte le fasi del ciclo di vita del sistema di IA, oltre che adeguatamente capaci di gestire risultati sbagliati. I sistemi di IA devono essere **affidabili**, sufficientemente sicuri da essere **resilienti** sia agli attacchi palesi sia a tentativi più subdoli di manipolazione dei dati o degli algoritmi, e devono garantire l'esistenza di un **piano di emergenza** in caso di problemi. Le loro decisioni devono essere **accurate**, o almeno rispecchiare correttamente il loro livello di accuratezza, e i loro risultati devono essere **riproducibili**.

I sistemi di IA dovrebbero inoltre contenere meccanismi di sicurezza fin dalla progettazione, per garantire che siano **sicuri in modo verificabile** in ogni fase, considerando soprattutto la sicurezza fisica e mentale di tutte le persone coinvolte. Ciò comprende anche la possibilità di ridurre al minimo e, ove possibile, rendere reversibili gli effetti involontari o gli errori del funzionamento del sistema. È opportuno prevedere processi in grado di chiarire e valutare i potenziali rischi associati all'uso dei sistemi di IA nei vari settori di applicazione.

III. Riservatezza e governance dei dati

La tutela della riservatezza e la **protezione dei dati** devono essere garantite **in tutte le fasi** del ciclo di vita del sistema di IA. Le registrazioni digitali del comportamento umano possono consentire ai sistemi di IA di dedurre non solo le preferenze, l'età e il genere dei singoli, ma anche il loro orientamento sessuale e il loro credo religioso o politico. Per far sì che le persone abbiano fiducia nel trattamento dei dati occorre garantire loro il pieno controllo dei dati personali e assicurare che i dati che li riguardano non saranno utilizzati per arrecare danno o discriminarli.

Oltre alla salvaguardia della riservatezza e dei dati personali, devono essere soddisfatti i requisiti necessari per garantire che i sistemi di IA siano di qualità elevata. La qualità dei set di dati utilizzati è fondamentale per le prestazioni dei sistemi di IA. Quando si raccolgono dati, questi possono riflettere condizionamenti di tipo sociale o contenere inesattezze, errori e vizi materiali. Questo aspetto deve essere risolto prima di utilizzare

¹² Il regolamento generale sulla protezione dei dati conferisce ai singoli il diritto di non essere sottoposti a una decisione basata unicamente sul trattamento automatizzato quando questa produca effetti giuridici che li riguardano o incida in modo analogo significativamente sulla loro persona (articolo 22 di tale regolamento).

¹³ Per "human-in-the-loop" (HITL) si intende l'intervento umano in ogni ciclo decisionale del sistema, cosa che in molti casi non è né possibile né auspicabile. Per "human-on-the-loop" (HOTL) si intende la capacità di intervento umano durante il ciclo di progettazione del sistema e di monitoraggio del funzionamento del sistema. Per "human-in-command" (HIC) si intende sia la capacità di sorvegliare l'attività complessiva del sistema di IA (compreso il più ampio impatto economico, sociale, giuridico ed etico) sia la capacità di decidere quando e come utilizzare il sistema in una particolare situazione. Si può per esempio decidere di non utilizzare un sistema di IA in una particolare situazione, di definire livelli di discrezionalità umana durante l'uso del sistema o di garantire la possibilità di annullare una decisione adottata dal sistema.

una qualsiasi set di dati per addestrare un sistema di IA. Deve essere inoltre garantita l'**integrità** dei dati. I processi e i set di dati utilizzati devono essere testati e documentati in ogni fase, come la pianificazione, l'addestramento, i test e la diffusione. Ciò dovrebbe valere anche per i sistemi di IA che non sono stati sviluppati internamente ma acquisiti altrove. Deve essere infine adeguatamente disciplinato e controllato l'**accesso** ai dati.

IV. Trasparenza

La **tracciabilità** dei sistemi di IA dovrebbe essere garantita: è importante registrare e documentare sia le decisioni adottate dai sistemi sia l'intero processo che ha prodotto le decisioni, comprese una descrizione della raccolta e dell'etichettatura dei dati e una descrizione dell'algoritmo utilizzato. In questo contesto dovrebbe essere prevista, per quanto possibile, la **spiegabilità** del processo decisionale degli algoritmi, adattata alle persone coinvolte. È opportuno proseguire la ricerca in corso volta a sviluppare meccanismi di spiegabilità. Dovrebbero anche essere disponibili spiegazioni sulla misura in cui un sistema di IA influenza e definisce il processo decisionale organizzativo, le scelte di progettazione del sistema e la logica alla base della sua diffusione, in modo da garantire la trasparenza non solo dei dati e dei sistemi, ma anche dei modelli di business.

Infine, è importante **comunicare**, opportunamente e in modo adeguato al caso in esame, le capacità e i limiti del sistema di IA ai diversi portatori di interessi coinvolti. I sistemi di IA dovrebbero inoltre essere identificabili come tali, così che gli utenti sappiano che stanno interagendo con un sistema di IA e possano individuare le persone che ne sono responsabili.

V. Diversità, non discriminazione ed equità

I set di dati utilizzati dai sistemi di IA (sia per l'addestramento sia per il funzionamento) possono essere inficiati da condizionamenti storici involontari, incompletezza e modelli di governance inadatti. Se mantenuti, tali condizionamenti potrebbero portare a discriminazioni (in)dirette. Un danno può derivare anche dallo sfruttamento intenzionale dei condizionamenti (dei consumatori) o da una concorrenza sleale. Può inoltre essere inficiato da condizionamenti anche il modo in cui sono sviluppati i sistemi di IA, ad esempio il modo in cui è scritto il codice di programmazione di un algoritmo. Tali preoccupazioni dovrebbero essere affrontate già dalle prime fasi dello sviluppo del sistema.

Per superarle possono essere utili anche la creazione di **gruppi di progettazione diversificati** e l'istituzione di meccanismi per garantire la **partecipazione**, in particolare dei cittadini, allo sviluppo dell'IA. È consigliabile consultare i portatori di interessi che possono essere direttamente o indirettamente interessati dal sistema durante il suo ciclo di vita. I sistemi di IA dovrebbero tenere in considerazione l'intera gamma delle capacità, delle competenze e dei bisogni umani e garantire l'accessibilità attraverso un approccio di progettazione universale che si sforzi di ottenere la parità di accesso per le persone con disabilità.

VI. Benessere sociale e ambientale

Per ottenere un'IA affidabile si dovrebbe tenere conto del suo impatto sull'**ambiente e sugli altri esseri senzienti**. Idealmente tutti gli esseri umani, comprese le generazioni

future, devono poter beneficiare della biodiversità e di un ambiente abitabile. La sostenibilità e la **responsabilità ecologica** dei sistemi di IA dovrebbero pertanto essere incoraggiate. Lo stesso vale per le soluzioni di IA che rispondono a criticità di portata mondiale, come ad esempio gli obiettivi di sviluppo sostenibile delle Nazioni Unite.

L'impatto dei sistemi di IA dovrebbe inoltre essere considerato non solo da una prospettiva individuale, ma anche dal punto di vista della **società nel suo complesso**. Andrebbe prestata particolare attenzione all'uso dei sistemi di IA soprattutto nelle situazioni che riguardano il processo democratico, compresi la formazione di opinioni, i processi decisionali politici o i contesti elettorali. Va inoltre tenuto in considerazione l'**impatto sociale** dell'IA. I sistemi di IA possono essere utilizzati per migliorare le competenze sociali, ma possono anche contribuire al loro deterioramento.

VII. Accountability.

Dovrebbero essere previsti meccanismi che garantiscano la responsabilità e l'accountability dei sistemi di IA e dei loro risultati, sia prima che dopo la loro attuazione. La **verificabilità** dei sistemi di IA è fondamentale in questo contesto, in quanto la valutazione dei sistemi di IA da parte di revisori interni ed esterni e la disponibilità di tali relazioni di valutazione contribuiscono fortemente all'affidabilità della tecnologia. La verificabilità esterna dovrebbe essere garantita soprattutto per le applicazioni che incidono sui diritti fondamentali, comprese quelle critiche per la sicurezza.

I potenziali impatti negativi dei sistemi di IA dovrebbero essere individuati, valutati, documentati e ridotti al minimo. Questo processo è facilitato dall'impiego delle valutazioni d'impatto, che dovrebbero essere proporzionate all'entità dei rischi che i sistemi di IA comportano. Le **contrapposizioni** tra i requisiti, che spesso sono inevitabili, dovrebbero essere affrontate in modo razionale e metodologico e se ne dovrebbe tenere conto. Infine, in caso di ingiusto impatto negativo, dovrebbero essere previsti meccanismi accessibili in grado di garantire **adeguati mezzi di riparazione**.

2.3. Prossime tappe: una fase pilota che coinvolga il più ampiamente possibile i portatori di interessi

Raggiungere un consenso su questi requisiti fondamentali per i sistemi di IA è una prima tappa importante verso la definizione di orientamenti per un'IA etica. Come prossimo passo, la Commissione garantirà che tali orientamenti possano essere testati e attuati nella pratica.

A tal fine, la Commissione avvierà presto una fase pilota mirata che servirà a raccogliere un feedback strutturato dai portatori di interessi. Questo esercizio si concentrerà in particolare sulla lista di controllo che il gruppo di esperti ad alto livello ha elaborato per ciascuno dei requisiti fondamentali.

Questo esercizio si articolerà in due parti: i) una fase pilota per gli orientamenti che coinvolge i portatori di interessi che sviluppano o utilizzano l'IA, comprese le amministrazioni pubbliche, e ii) una consultazione continua dei portatori di interessi e un processo di sensibilizzazione in tutti gli Stati membri e per i diversi gruppi di portatori di interessi, compresi i settori dell'industria e dei servizi:

- i) a partire dal giugno 2019 tutti i portatori di interessi e tutti i cittadini saranno invitati a testare la lista di controllo e a fornire un feedback su come migliorarla. Il gruppo di esperti ad alto livello sull'IA avvierà inoltre un riesame approfondito con i portatori

di interessi del settore pubblico e privato per raccogliere un feedback più dettagliato su come attuare gli orientamenti in un'ampia gamma di settori di applicazione. Il feedback ricevuto sulla praticabilità e sulla fattibilità degli orientamenti sarà valutato entro la fine del 2019;

- ii) parallelamente la Commissione organizzerà ulteriori attività di comunicazione che daranno ai rappresentanti del gruppo di esperti ad alto livello sull'IA l'opportunità di presentare gli orientamenti ai portatori di interessi degli Stati membri, compresi i settori dell'industria e dei servizi, e di fornire loro un'ulteriore opportunità di formulare osservazioni in merito agli orientamenti sull'IA e contribuire agli stessi.

La Commissione terrà conto del lavoro del gruppo di esperti sull'etica per quanto riguarda la guida connessa e automatizzata¹⁴ e opererà in collaborazione con progetti di ricerca finanziati dall'UE nel campo dell'IA e con i pertinenti partenariati pubblico-privato per l'attuazione dei requisiti fondamentali¹⁵. Ad esempio la Commissione sosterrà, in coordinamento con gli Stati membri, lo sviluppo di una banca dati comune di immagini medicali, inizialmente dedicata alle forme più comuni di tumore, che consentirà di addestrare gli algoritmi a diagnosticare i sintomi con una precisione molto elevata. Analogamente, la collaborazione tra la Commissione e gli Stati membri permetterà di realizzare un numero sempre maggiore di corridoi transfrontalieri per le prove dei veicoli connessi e automatizzati. Gli orientamenti dovrebbero essere applicati e testati in questi progetti e i risultati saranno integrati nel processo di valutazione.

Per la fase pilota e la consultazione dei portatori di interessi si potrà contare sul contributo dell'Alleanza europea per l'IA e di AI4EU, la piattaforma di IA "on demand". Il progetto AI4EU¹⁶, avviato nel gennaio 2019, riunisce algoritmi, strumenti, set di dati e servizi per aiutare le organizzazioni, in particolare le piccole e medie imprese, ad attuare soluzioni di IA. L'Alleanza europea per l'IA, insieme a AI4EU, continuerà a mobilitare l'ecosistema dell'IA in tutta Europa, anche nel contesto della fase pilota degli orientamenti etici sull'IA e per promuovere il rispetto per un'IA antropocentrica.

All'inizio del 2020, sulla base della valutazione del feedback ricevuto durante la fase pilota, il gruppo di esperti ad alto livello sull'IA riesaminerà e aggiornerà gli orientamenti. Analizzati il riesame e l'esperienza acquisita, la Commissione valuterà i risultati e proporrà eventuali passi successivi.

L'IA etica è una proposta vantaggiosa per tutti. Garantire il rispetto dei valori e dei diritti fondamentali non solo è un elemento essenziale di per sé, ma facilita anche l'accettazione da parte del pubblico e aumenta il vantaggio competitivo delle imprese europee attive nel campo dell'IA creando un'IA antropocentrica e affidabile, nota per i suoi prodotti etici e sicuri e, più in generale, forte della solida reputazione delle imprese europee nel fornire prodotti sicuri e di alta qualità. La fase pilota contribuirà a garantire che i prodotti dell'IA mantengano tale promessa.

2.4. Verso orientamenti etici internazionali sull'IA

¹⁴ Cfr. la comunicazione della Commissione sulla mobilità connessa e automatizzata, COM(2018) 283.

¹⁵ Nel quadro del Fondo europeo per la difesa la Commissione elaborerà inoltre orientamenti etici specifici per la valutazione delle proposte di progetto nel settore dell'IA per la difesa.

¹⁶ <https://ec.europa.eu/digital-single-market/en/news/artificial-intelligence-ai4eu-project-launches-1-january-2019>.

Le discussioni internazionali sull'etica dell'IA si sono intensificate dopo che la presidenza giapponese del G7 ha posto tale tema tra le priorità all'ordine del giorno nel 2016. Date le interconnessioni internazionali dello sviluppo dell'IA in termini di circolazione di dati, sviluppo di algoritmi e investimenti nella ricerca, **la Commissione continuerà ad adoperarsi per portare l'approccio dell'Unione sulla scena mondiale e costruire un consenso su un'IA antropocentrica¹⁷.**

Il lavoro svolto dal gruppo di esperti ad alto livello sull'IA, in particolare l'elenco dei requisiti e il processo di impegno con i portatori di interessi, fornisce alla Commissione un ulteriore contributo prezioso da apportare alle discussioni internazionali. L'Unione europea può svolgere un ruolo guida nello sviluppo di orientamenti internazionali sull'IA e, se possibile, di un meccanismo di valutazione correlato.

La Commissione intende pertanto:

rafforzare la cooperazione con i partner che condividono gli stessi principi:

- valutando in che misura si possa giungere a una convergenza con i progetti di orientamenti etici di paesi terzi (ad esempio Giappone, Canada, Singapore) e, assieme a questo gruppo di paesi che condividono gli stessi principi, preparare una discussione più ampia, sostenuta da azioni di attuazione dello strumento di partenariato per la cooperazione con i paesi terzi¹⁸ e
- valutare in che modo le imprese di paesi terzi e le organizzazioni internazionali possano contribuire alla "fase pilota" degli orientamenti testandoli e convalidandoli;

continuare a svolgere un ruolo attivo nelle discussioni e nelle iniziative internazionali:

- contribuire ai consessi multilaterali quali il G7 e il G20;
- avviare dialoghi con paesi terzi e organizzare riunioni bilaterali e multilaterali per creare un consenso sull'IA antropocentrica;
- contribuire alle attività di normazione pertinenti nell'ambito degli organismi internazionali di normazione per promuovere questa visione e
- rafforzare la raccolta e la diffusione delle conoscenze sulle politiche pubbliche, collaborando con le organizzazioni internazionali competenti.

¹⁷ L'alto rappresentante dell'Unione per gli affari esteri e la politica di sicurezza, con il sostegno della Commissione, si baserà sulle consultazioni in sede di Nazioni Unite, Global Tech Panel e altri forum multilaterali, e in particolare coordinerà le proposte per affrontare le sfide complesse in gioco in tema di sicurezza.

¹⁸ Regolamento (UE) n. 234/2014 del Parlamento europeo e del Consiglio, dell'11 marzo 2014, che istituisce uno strumento di partenariato per la cooperazione con i paesi terzi (GU L 77 del 15.3.2014, pag. 77). Ad esempio, il previsto progetto "Un'alleanza internazionale per un approccio antropocentrico all'intelligenza artificiale" faciliterà la realizzazione di iniziative congiunte con partner che condividono gli stessi principi al fine di promuovere orientamenti etici e adottare principi comuni e conclusioni operative. Consentirà all'UE e a tali paesi di discutere le conclusioni operative che derivano dagli orientamenti etici sull'IA proposti dal gruppo di esperti al fine di raggiungere un approccio comune. Permetterà inoltre di monitorare l'adozione della tecnologia dell'IA in tutto il mondo. Infine, il progetto prevede l'organizzazione di attività di diplomazia pubblica che accompagnino eventi internazionali, ad esempio a livello del G7, del G20 e dell'Organizzazione per la cooperazione e lo sviluppo economico.

3. CONCLUSIONI

L'UE si fonda su una serie di valori fondamentali e su queste fondamenta ha costruito un quadro normativo solido ed equilibrato. Sulla base di questo quadro normativo esistente, si avverte la necessità di elaborare orientamenti etici per lo sviluppo e l'uso dell'IA in ragione della sua novità e delle sfide specifiche che questa tecnologia comporta. L'IA può essere considerata affidabile solo se è sviluppata e utilizzata nel rispetto di valori etici ampiamente condivisi.

In considerazione di tale obiettivo, la Commissione accoglie con favore il contributo preparato dal gruppo di esperti ad alto livello sull'IA. Sulla base dei requisiti fondamentali perché l'IA sia considerata affidabile, la Commissione avvierà ora una fase pilota mirata per garantire che gli orientamenti etici per lo sviluppo e l'utilizzo dell'IA che ne derivano possano essere attuati nella pratica. La Commissione si adopererà inoltre per creare un ampio consenso sociale sull'IA antropocentrica coinvolgendo tutti portatori di interessi e i partner internazionali.

La dimensione etica dell'IA non è un'opzione di lusso né un complemento accessorio, è necessario invece che sia parte integrante dello sviluppo dell'IA. Puntando a un'IA antropocentrica e basata sulla fiducia salvaguardiamo il rispetto dei nostri valori sociali fondamentali e consentiamo all'Europa e alla sua industria di distinguersi chiaramente in quanto leader di un'IA all'avanguardia affidabile in tutto il mondo.

Per garantire lo sviluppo etico dell'IA in Europa nel suo contesto più ampio, la Commissione sta perseguendo un approccio globale che comprende in particolare le seguenti linee di azione da attuare entro il terzo trimestre del 2019:

- avvierà una serie di **reti di centri di eccellenza per la ricerca sull'IA** tramite Orizzonte 2020. Selezionerà un massimo di quattro reti, incentrate in particolare sulle maggiori sfide scientifiche o tecnologiche come la spiegabilità e l'interazione avanzata uomo-macchina, elementi fondamentali per un'IA affidabile;
- avvierà la creazione di **reti di poli dell'innovazione digitale**¹⁹ incentrati sull'IA per le attività produttive e sui big data;
- insieme agli Stati membri e ai portatori di interessi la Commissione avvierà discussioni preparatorie per sviluppare e attuare un **modello per la condivisione dei dati e per utilizzare al meglio gli spazi di dati comuni**, in particolare nei settori dei trasporti, dell'assistenza sanitaria e della produzione industriale²⁰.

La Commissione sta inoltre lavorando a una relazione sulle sfide che l'IA pone per i quadri normativi in materia di responsabilità e sicurezza e a un documento di orientamento sull'attuazione della direttiva sulla responsabilità per danno da prodotti difettosi²¹. In contemporanea l'iniziativa europea per il calcolo ad alte prestazioni (EuroHPC)²² svilupperà la prossima generazione di supercomputer, in quanto la capacità di calcolo è essenziale per il trattamento dei dati e l'addestramento dell'IA e l'Europa deve padroneggiare l'intera catena del

¹⁹ <http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs>.

²⁰ Le risorse necessarie saranno mobilitate da Orizzonte 2020 (che prevede una dotazione di quasi 1,5 miliardi di EUR per l'IA per il periodo 2018-2020) e dal suo previsto successore, Orizzonte Europa, dalla componente digitale del meccanismo per collegare l'Europa e soprattutto dal futuro programma Europa digitale. I progetti saranno anche finanziati da risorse provenienti dal settore privato e da programmi degli Stati membri.

²¹ Cfr. la comunicazione della Commissione "L'intelligenza artificiale per l'Europa", COM(2018) 237.

²² <https://eurohpc-ju.europa.eu>.

valore digitale. Il partenariato in corso con gli Stati membri e con l'industria su componenti e sistemi microelettronici (ECSEL)²³ e l'iniziativa europea in materia di processori²⁴ contribuiranno allo sviluppo della tecnologia per processori a bassa potenza per un calcolo ad alte prestazioni e un edge computing affidabili e sicuri.

Come l'attività sugli orientamenti etici per l'IA, anche queste iniziative si basano su una **stretta collaborazione tra tutti i portatori di interessi**, gli Stati membri, l'industria, gli attori sociali e i cittadini. Nel complesso l'approccio dell'Europa all'intelligenza artificiale dimostra che la competitività economica e la fiducia della società devono partire dagli stessi valori fondamentali e rafforzarsi reciprocamente.

²³ www.ecsel.eu.

²⁴ www.european-processor-initiative.eu.