EUROPEISKA
KOMMISSIONEN

Bryssel den 8.4.2019
COM(2019) 168 final

MEDDELANDE FRÅN KOMMISSIONEN TILL EUROPAPARLAMENTET, RÅDET, EUROPEISKA EKONOMISKA OCH SOCIALA KOMMITTÉN SAMT REGIONKOMMITTÉN

Att skapa förtroende för människocentrerad artificiell intelligens

1. INLEDNING – DEN EUROPEISKA AI-STRATEGIN

Artificiell intelligens (AI) har potential att förändra vår värld till det bättre: den kan förbättra hälso- och sjukvården, minska energiförbrukningen, göra bilarna säkrare och ge jordbrukare möjligheter att utnyttja vatten och naturresurser på ett ändamålsenligare sätt. AI kan användas för att förutsäga miljö- och klimatförändringar och förbättra den finansiella riskhanteringen, och gör det möjligt att med mindre avfall tillverka produkter som är anpassade till våra behov. AI kan också bidra till att bedrägerier och cybersäkerhetshot upptäcks, och hjälper brottsbekämpande myndigheter att effektivare bekämpa brottslighet.

Den artificiella intelligensen kan gynna hela samhället och ekonomin. Den är en strategisk teknik som just nu håller på att utvecklas och tas i bruk i rask takt runtom i världen. AI medför emellertid även nya utmaningar för sysselsättningens framtid och ger upphov till rättsliga och etiska frågor.

För att bemöta dessa utmaningar och tillvarata alla möjligheter med artificiell intelligens offentliggjorde kommissionen en europeisk strategi[1] **i april 2018.** I strategin placeras människorna i centrum för utvecklingen av AI – **människocentrerad AI**. Den är en tredelad strategi för att stärka EU:s tekniska och industriella kapacitet och spridningen av AI i hela ekonomin, lägga grunden för socioekonomiska förändringar och sörja för en lämplig etisk och rättslig ram.

I syfte att förverkliga AI-strategin har **kommissionen tillsammans med medlemsstaterna utvecklat en samordnad AI-plan**[2], som presenterades i december 2018, för att skapa synergier, samla data – råmaterialet för många AI-tillämpningar – och öka de gemensamma investeringarna. Målet är att främja gränsöverskridande samarbete och mobilisera alla aktörer för att öka de offentliga och privata investeringarna till **minst 20 miljarder euro** årligen under de kommande tio åren[3]. Kommissionen fördubblade sina investeringar i AI i Horisont 2020 och planerar att investera 1 miljard euro årligen från Horisont Europa och programmet för ett digitalt Europa, framför allt för att stödja gemensamma dataområden inom hälsa, transport och tillverkning samt stora experimentanläggningar såsom smarta sjukhus och infrastrukturer för automatiserade fordon och en strategisk forskningsagenda.

För att genomföra en sådan gemensam strategisk agenda för forskning, innovation och spridning har kommissionen intensifierat sin **dialog med alla berörda parter** från näringslivet, forskningsinstitut och myndigheter. Det nya programmet för ett digitalt Europa kommer också att vara avgörande för att göra AI tillgänglig för små och medelstora företag i alla medlemsstater genom digitala innovationsknutpunkter, utökade testnings- och experimentanläggningar, dataområden och utbildningsprogram.

EU:s etiska förhållningssätt till artificiell intelligens utgår från Europas anseende för säkra produkter av hög kvalitet för att stärka allmänhetens förtroende för den digitala utvecklingen och skapa en konkurrensfördel för europeiska AI-företag. Syftet med detta meddelande är att inleda en övergripande pilotfas med så många olika berörda parter som möjligt för att testa det praktiska genomförandet av den etiska vägledningen för utveckling och användning av AI.

2. ATT SKAPA FÖRTROENDE FÖR MÄNNISKOCENTRERAD AI

---

1 COM(2018) 237.

2 COM(2018) 795.

3 För att uppnå detta mål har kommissionen föreslagit att unionen under nästa programperiod 2021–2027 anslår minst 1 miljard euro per år i finansiering från Horisont Europa och programmet för ett digitalt Europa för investeringar i AI.

EU:s AI-strategi och samordnade plan klargör att **förtroende är en förutsättning för ett människocentrerat förhållningssätt till artificiell intelligens** – AI är inte ett självändamål utan ett verktyg som ska hjälpa människor, med det yttersta målet att öka människornas välbefinnande. För att uppnå detta bör man **se till att den artificiella intelligensen är tillförlitlig**. De värden som våra samhällen bygger på måste integreras helt och hållet i utvecklingen av AI.

Unionen bygger på **värdena respekt för människans värdighet, frihet, demokrati, jämlikhet, rättsstaten och respekt för de mänskliga rättigheterna**, inklusive rättigheter för personer som tillhör minoriteter[4]. Dessa värden är gemensamma för medlemsstaterna i ett samhälle som kännetecknas av mångfald, icke-diskriminering, tolerans, rättvisa, solidaritet och jämställdhet. Dessutom sammanställs i **EU:s stadga om de grundläggande rättigheterna** i en enda text de personliga, medborgerliga, politiska, ekonomiska och sociala rättigheter som människor i EU åtnjuter.

EU har ett **starkt regelverk** som kommer att utgöra den globala standarden för människocentrerad AI. Den allmänna dataskyddsförordningen garanterar en hög skyddsnivå för personuppgifter och kräver att åtgärder vidtas för att säkerställa inbyggt dataskydd och dataskydd som standard[5]. Förordningen om det fria flödet av uppgifter som inte är personuppgifter avlägsnar hinder för den fria rörligheten för sådana uppgifter och tryggar behandlingen av alla kategorier av uppgifter överallt i Europa. Den nyligen antagna cybersäkerhetsakten kommer att hjälpa till att stärka förtroendet för onlinevärlden, något som även den föreslagna förordningen om integritet och elektronisk kommunikation[6] syftar till.

Trots detta medför artificiell intelligens nya utmaningar eftersom den gör det möjligt för maskiner att ”lära sig” och att fatta och genomföra beslut utan mänsklig medverkan. Inom kort kommer denna typ av funktion att bli standard i många typer av varor och tjänster, allt från smarttelefoner till självkörande bilar, robotar och onlineapplikationer. Men beslut som fattas med hjälp av algoritmer kan vara ett resultat av data som är ofullständiga och därför inte tillförlitliga, de kan vara manipulerade genom cyberangrepp, de kan vara snedvridna eller helt enkelt felaktiga. Att urskillningslöst tillämpa tekniken medan den håller på att utvecklas skulle därför leda till problematiska resultat och att allmänheten blir tveksam till att godta den eller använda den.

I stället bör AI-tekniken utvecklas med människorna i centrum, så att den blir värd allmänhetens förtroende. Detta innebär att AI-tillämpningar inte bara bör vara förenliga med lagstiftningen, utan även följa etiska principer och se till att oavsiktliga skador undviks i praktiken. Mångfald i fråga om kön, ras eller etniskt ursprung, religion eller övertygelse, funktionsvariation och ålder bör garanteras i varje skede av AI-utvecklingen. AI-tillämpningar bör ge medborgarna egenmakt och respektera deras grundläggande rättigheter. Syftet bör vara att förbättra människors förmågor, inte att ersätta dem, och de bör även vara tillgängliga för personer med funktionsnedsättning.

Det finns alltså ett behov av **etiska riktlinjer** som bygger på den gällande rättsliga ramen och som bör tillämpas av utvecklare, leverantörer och användare av artificiell intelligens på den

---

4 EU är även part i FN:s konvention om rättigheter för personer med funktionsnedsättning.

5 Förordning (EU) 2016/679. Den allmänna dataskyddsförordningen garanterar det fria flödet av personuppgifter inom unionen och innehåller bestämmelser om beslutsfattande som grundas uteslutande på automatisk behandling, inklusive profilering. De personer som berörs har rätt att bli informerade om det automatiserade beslutsfattandet och att få meningsfull information om logiken i det, vikten av behandlingen och hur den kan tänkas påverka dem. De har också rätt att i sådana fall kräva mänsklig medverkan, att uttrycka sin åsikt och att bestrida beslutet.

6 COM(2017) 10.

inre marknaden, så att samma etiska villkor gäller i alla medlemsstater. Därför har kommissionen inrättat en **högnivåexpertgrupp för artificiell intelligens**[7] (nedan kallad *AI-expertgruppen*) med företrädare för många olika berörda parter och gett den i uppgift att utarbeta etiska riktlinjer för AI och rekommendationer för den mer övergripande AI-politiken. Samtidigt inrättades den **europeiska AI-alliansen**[8], en öppen flerpartsplattform med över 2 700 medlemmar, för att ge ett bredare bidrag till arbetet i AI-expertgruppen.

AI-expertgruppen offentliggjorde ett första utkast till etiska riktlinjer i december 2018. Efter ett **samråd med berörda parter**[9] och **möten med företrädare för medlemsstaterna**[10] lämnade AI-expertgruppen in ett reviderat dokument till kommissionen i mars 2019. Berörda parter har i sin återkoppling hittills på det hela taget välkomnat riktlinjernas praktiska inriktning och den konkreta vägledning de erbjuder utvecklare, leverantörer och användare av AI om hur man kan säkerställa tillförlitlighet.

2.1. AI-expertgruppens riktlinjer för tillförlitlig AI

De riktlinjer som utarbetats av AI-expertgruppen och som hänvisas till i detta meddelande[11] bygger i synnerhet på det arbete som gjorts inom Europeiska gruppen för etik inom vetenskap och ny teknik och Europeiska unionens byrå för grundläggande rättigheter.

Enligt riktlinjerna krävs tre komponenter för att artificiell intelligens ska vara ”tillförlitlig”: 1) den bör vara förenlig med lagstiftningen, 2) den bör uppfylla etiska principer och 3) den bör vara robust.

På grundval av dessa tre komponenter och de europeiska värden som anges i avsnitt 2 anges i riktlinjerna sju centrala krav som AI-tillämpningar bör uppfylla för att betraktas som tillförlitliga. Riktlinjerna innehåller också en bedömningslista som gör det lättare att kontrollera om dessa krav är uppfyllda.

De sju centrala kraven är följande:

- Mänskligt agentskap och mänsklig tillsyn.
- Teknisk robusthet och säkerhet.
- Integritet och dataförvaltning.
- Transparens.
- Mångfald, icke-diskriminering och rättvisa.
- Samhällets och miljöns välbefinnande.
- Ansvarsskyldighet.

---

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 Detta samråd ledde till kommentarer från 511 organisationer, föreningar, företag, forskningsinstitut, privatpersoner och andra. En sammanfattning av de inkomna synpunkterna finns på https://ec.europa.eu/futurium/en/system/files/ged/consultation_feedback_on_draft_ai_ethics_guidelines_4.pdf

10 AI-expertgruppens arbete har mottagits positivt av medlemsstaterna, och rådet noterade i sina slutsatser av den 18 februari 2019 bland annat det kommande offentliggörandet av de etiska riktlinjerna och stödde kommissionens ansträngningar för att etablera EU:s etiska strategi globalt: https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/sv/pdf

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

Även om dessa krav är tänkta att gälla alla AI-system inom olika miljöer och branscher bör det specifika sammanhang där de används beaktas för det konkreta och proportionella genomförandet, med en metod som grundar sig på inverkan. Till exempel är det mycket mindre farligt att en AI-tillämpning ger lästips om en ointressant bok än att den feldiagnosticerar cancer, och mindre sträng övervakning kan därför tillämpas.

AI-expertgruppens riktlinjer är inte bindande och skapar i sig inga nya rättsliga skyldigheter. Många befintliga (ofta användnings- eller domänspecifika) bestämmelser i unionsrätten återspeglar dock naturligtvis redan ett eller flera av dessa centrala krav, såsom regler om säkerhet, skydd av personuppgifter, integritet eller miljöskydd.

Kommissionen välkomnar AI-expertgruppens arbete och anser det vara ett värdefullt bidrag till utformningen av politiken.

2.2. Centrala krav för tillförlitlig AI

Kommissionen stöder följande centrala krav för tillförlitlig AI, som bygger på europeiska värden. Den uppmuntrar berörda parter att följa kraven och att testa listan för bedömning av deras praktiska tillämpning för att skapa rätt förtroendemiljö för en framgångsrik utveckling och användning av artificiell intelligens. Kommissionen välkomnar återkoppling från berörda parter för att utvärdera om denna bedömningslista som ingår i riktlinjerna kräver ytterligare justeringar.

I. <u>Mänskligt agentskap och mänsklig tillsyn</u>

AI-system bör hjälpa enskilda personer att göra bättre och mer välgrundade val i enlighet med sina mål. De bör fungera som katalysatorer för ett blomstrande och rättvist samhälle genom att stödja mänskligt agentskap och **de grundläggande rättigheterna**, och inte minska, begränsa eller snedvrida människans självständighet. **Användarens allmänna välbefinnande** bör vara av central betydelse för systemets funktionalitet.

Mänsklig tillsyn ser till att ett AI-system inte undergräver människors självständighet eller orsakar andra skadliga effekter. Passande nivåer av **kontrollåtgärder** bör vidtas[12] utifrån det specifika AI-baserade systemet och dess tillämpningsområde. Detta gäller bland annat AI-systemens anpassningsbarhet, korrekthet och förklarbarhet. **Tillsynen** kan uppnås genom styrmekanismer som t.ex. tillvägagångssätt med *human-in-the-loop*, *human-on-the-loop* eller *human-in-command*[13]. Det måste säkerställas att myndigheterna har möjlighet att utöva sina tillsynsbefogenheter i enlighet med sina mandat. Om förhållandena i övrigt är lika krävs mer omfattande testning och striktare styrning ju

[12] Den allmänna dataskyddsförordningen ger enskilda personer rätt att inte bli föremål för ett beslut som enbart grundas på automatiserad behandling när detta har rättsliga följder för användarna eller på liknande sätt i betydande grad påverkar dem (artikel 22 i den allmänna dataskyddsförordningen).

[13] *Human-in-the-loop* innebär att en människa medverkar i varje beslutscykel i systemet, vilket i många fall varken är möjligt eller önskvärt. *Human-on-the-loop* innebär att det finns en möjlighet till mänskligt ingripande under utformningen av systemet och övervakningen av systemets funktion. *Human-in-command* innebär en möjlighet att övervaka AI-systemets övergripande aktivitet (även dess vidare ekonomiska, samhälleliga, rättsliga och etiska inverkan) och förmåga att bestämma när och hur systemet ska användas i en viss situation. Det kan även innebära beslut om att inte använda ett AI-system i en viss situation, att inrätta nivåer av mänsklig handlingsfrihet under användningen av systemet eller att säkerställa möjligheten att kringgå ett beslut som fattas av systemet.

mindre tillsyn en människa kan utöva över ett AI-system.

II. Teknisk robusthet och säkerhet

Tillförlitlig AI kräver att algoritmerna är tillräckligt säkra, pålitliga och robusta för att klara av fel eller inkonsekvenser under alla faser av AI-systemets livscykel och att hantera felaktiga resultat på ett tillfredsställande sätt. AI-systemen måste vara **pålitliga**, tillräckligt säkra för att vara **motståndskraftiga** mot såväl uppenbara angrepp som mer subtila försök att manipulera data eller själva algoritmerna, och det måste finnas en **reservplan** i händelse av problem. Deras beslut måste vara **korrekta**, eller åtminstone rätt återspegla deras korrekthetsnivå, och resultaten bör vara **reproducerbara**.

Dessutom bör AI-system integrera mekanismer för trygghet och inbyggd säkerhet så att de är **garanterat säkra** i varje steg, med fokus på alla berörda parters fysiska och psykiska säkerhet. Till detta hör även att se till att oavsiktliga konsekvenser eller fel i systemdriften kan minimeras och om möjligt repareras. Processer bör införas för att klargöra och bedöma de potentiella riskerna med användningen av AI-system inom olika tillämpningsområden.

III. Integritet och dataförvaltning

Integritet och **dataskydd** måste garanteras i **alla skeden** av AI-systemets livscykel. Digital registrering av mänskligt beteende kan göra det möjligt för AI-system att inte bara sluta sig till personers preferenser, ålder och kön, utan även deras sexuella läggning, religion eller politiska åsikter. För att människor ska kunna lita på databehandlingen måste det säkerställas att de har full kontroll över sina egna data och att data om dem inte kommer att användas för att skada dem eller diskriminera dem.

Utöver att skydda integriteten och personuppgifter måste AI-systemen uppfylla höga kvalitetskrav. Kvaliteten på de data som används är avgörande för att AI-system ska fungera väl. När data samlas in kan de avspegla socialt konstruerad snedvridning eller innehålla oriktigheter, fel och misstag. Detta måste åtgärdas innan man tränar upp ett AI-system med ett givet dataset. Dessutom måste uppgifternas **integritet** garanteras. De processer och data som används måste testas och dokumenteras i varje steg, såsom planering, uppträning, testning och ibruktagande. Detta bör även gälla AI-system som inte har utvecklats internt utan införskaffats externt. Slutligen krävs tillräcklig reglering och kontroll av **åtkomsten** till data.

IV. Transparens

AI-systemens **spårbarhet** måste garanteras. Det är viktigt att logga och dokumentera både de beslut som fattas av systemen och hela den process som leder fram till besluten (inklusive en beskrivning av datainsamling och märkning samt en beskrivning av den algoritm som använts). I anslutning till detta bör man i möjlig mån se till att den algoritmbaserade beslutsprocessen kan **förklaras** på ett sätt som är anpassat till de berörda personerna. Pågående forskning för att utveckla förklaringsmekanismer bör fortsätta. Dessutom bör det finnas förklaringar av i vilken grad ett AI-system påverkar och formar den organisatoriska beslutsprocessen, systemets designval och orsakerna till att använda det (vilket därmed säkerställer inte bara data- och systemtransparens, utan även transparens kring verksamhetsmodellen).

Slutligen är det viktigt att på ett adekvat sätt **förmedla** AI-systemets kapacitet och begränsningar till de olika berörda parterna på ett sätt som lämpar sig för det aktuella fallet. Dessutom bör AI-system kunna identifieras som sådana, så att användarna vet att de interagerar med ett AI-system och vilka personer som ansvarar för det.

V. Mångfald, icke-diskriminering och rättvisa

Dataset som utnyttjas av AI-system (både när det tränas upp och används) kan vara påverkade av oavsiktlig snedvridning sedan tidigare, vara ofullständiga och följa dåliga styrmodeller. Om denna snedvridning fortsätter kan den leda till (in)direkt diskriminering. Skada kan också uppstå till följd av ett avsiktligt utnyttjande av (konsument)snedvridning eller genom illojal konkurrens. Dessutom kan AI-system även vara utvecklade på ett snedvridet sätt (t.ex. hur programmeringskoden för en algoritm är skriven). Sådana farhågor bör tas upp redan från början av systemutvecklingen.

De kan också tillmötesgås genom att man skapar **designteam med stor mångfald** och inrättar mekanismer som sörjer för att framför allt privatpersoner **deltar** i AI-utvecklingen. Det är lämpligt att samråda med berörda parter som direkt eller indirekt kan påverkas av systemet under hela dess livscykel. AI-system bör beakta hela skalan av mänskliga förmågor, färdigheter och krav, och säkerställa tillgänglighet genom en universell designstrategi som syftar till att uppnå lika tillgång för personer med funktionsnedsättningar.

VI. Samhällets och miljöns välbefinnande

För att artificiell intelligens ska vara tillförlitlig bör dess inverkan på **miljön och andra kännande varelser** beaktas. Idealet vore att alla människor, även kommande generationer, får leva i en mångfaldig och beboelig miljö. AI-systemens hållbarhet och **ekologiska ansvar** bör därför uppmuntras. Detsamma gäller AI-lösningar på områden av global betydelse, t.ex. FN:s mål för hållbar utveckling.

Därutöver bör man inte bara beakta hur AI-systemen påverkar enskilda, utan även hur de påverkar **samhället i stort**. Användningen av AI-system bör övervägas noggrant i synnerhet i situationer som rör den demokratiska processen, bland annat opinionsbildning, politiskt beslutsfattande eller valsammanhang. Beaktas bör även den artificiella intelligensens **sociala inverkan**. AI-system kan visserligen användas för att öka de sociala färdigheterna, men de kan också bidra till att de försämras.

VII. Ansvarsskyldighet

Mekanismer bör inrättas för att säkerställa ansvarighet och ansvarsskyldighet för AI-system och deras resultat, både före och efter det att de tas i bruk. I detta sammanhang är det viktigt att AI-systemen kan **granskas**, eftersom teknikens tillförlitlighet i stor utsträckning påverkas av interna och externa granskares bedömning av AI-system och tillgången till sådana utvärderingsrapporter. I synnerhet bör man se till att extern granskning är möjlig för tillämpningar som påverkar de grundläggande rättigheterna, t.ex. säkerhetskritiska tillämpningar.

AI-systemens **potentiella negativa effekter** bör identifieras, bedömas, dokumenteras och minimeras. Användningen av konsekvensbedömningar underlättar denna process. Dessa bedömningar bör stå i proportion till omfattningen av de risker som AI-systemen utgör.

Kompromisser mellan kraven – som ofta är oundvikliga – bör hanteras på ett rationellt och metodiskt sätt och bör redovisas. Slutligen bör det finnas tillgängliga mekanismer som garanterar **lämplig prövning** om orättvisa negativa konsekvenser uppstår.

2.3. Nästa steg: en pilotfas med många olika berörda parter

En första viktig milstolpe på vägen mot riktlinjer för etisk artificiell intelligens är att nå samförstånd om dessa centrala krav för AI-system. Som ett nästa steg kommer kommissionen att se till att denna vägledning kan testas och genomföras i praktiken.

I detta syfte kommer kommissionen nu att inleda en riktad pilotfas som är utformad för att få strukturerad återkoppling från berörda parter. Denna övning kommer i synnerhet att fokusera på den bedömningslista som AI-expertgruppen har utarbetat för vart och ett av de centrala kraven.

Arbete kommer att bestå av två delar: i) en pilotfas för riktlinjerna med berörda parter som utvecklar eller använder AI, även offentliga förvaltningar, och ii) fortsatta samråd med berörda parter och medvetandehöjande åtgärder i medlemsstaterna och bland olika grupper av berörda parter, bland annat industri- och tjänstesektorerna.

(i) Från och med juni 2019 kommer alla berörda parter och enskilda personer att uppmanas att testa bedömningslistan och ge återkoppling om hur den kan förbättras. Dessutom kommer AI-expertgruppen att göra en ingående granskning med berörda parter från den privata och den offentliga sektorn för att samla in mer detaljerad återkoppling om hur riktlinjerna kan genomföras inom ett stort antal tillämpningsområden. All återkoppling om riktlinjernas genomförbarhet och rimlighet kommer att utvärderas senast i slutet av 2019.

(ii) Parallellt med detta kommer kommissionen att anordna ytterligare uppsökande verksamhet och ge företrädare för AI-expertgruppen möjlighet att presentera riktlinjerna för berörda parter i medlemsstaterna, bland annat inom industri- och tjänstesektorerna. På så sätt får dessa berörda parter ytterligare en möjlighet att kommentera och bidra till riktlinjerna.

Kommissionen kommer att beakta arbetet i gruppen av experter på etiska frågor vid uppkopplad och automatiserad körning[14] och arbeta med EU-finansierade forskningsprojekt kring AI och med relevanta offentlig-privata partnerskap kring genomförandet av de centrala kraven[15]. Kommissionen kommer t.ex. att i samarbete med medlemsstaterna stödja utvecklingen av en gemensam databas med hälsobilder som inledningsvis inriktas på de vanligaste cancerformerna, så att algoritmer kan tränas upp att med mycket hög noggrannhet diagnostisera symtom. I likhet med detta möjliggör samarbetet mellan kommissionen och medlemsstaterna ett ökande antal gränsöverskridande utrymmen för testning av uppkopplade och automatiserade fordon. Riktlinjerna bör tillämpas och testas vid dessa projekt, och resultaten kommer att användas i utvärderingsprocessen.

Pilotfasen och samrådet med berörda parter kommer att gynnas av bidraget från den europeiska AI-alliansen och AI4EU, en AI on demand-plattform. Projektet AI4EU[16], som lanserades i januari 2019, sammanför algoritmer, verktyg, dataset och tjänster för att hjälpa

[14] Se kommissionens meddelande om uppkopplad och automatiserad rörlighet, COM(2018) 283.

[15] Inom ramen för Europeiska försvarsfonden kommer kommissionen också att utarbeta särskilda etiska riktlinjer för utvärdering av projektförslag på området artificiell intelligens inom försvaret.

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

organisationer, i synnerhet små och medelstora företag, att ta i bruk AI-lösningar. Den europeiska AI-alliansen kommer tillsammans med AI4EU att fortsätta att mobilisera AI-ekosystem i hela Europa, även i syfte att testa de etiska riktlinjerna för AI och främja respekten för människocentrerad AI.

I början av 2020 kommer **AI-expertgruppen att se över och uppdatera riktlinjerna**, utgående från utvärderingen av den återkoppling som inkommit under pilotfasen. På grundval av översynen och de erfarenheter som gjorts kommer **kommissionen att utvärdera resultatet och föreslå nästa steg**.

Etisk AI är ett alternativ som alla vinner på. Att garantera respekten för de grundläggande värdena och rättigheterna är viktigt i sig, men det underlättar också allmänhetens acceptans och ökar de europeiska AI-företagens konkurrensfördelar genom att man etablerar ett varumärke av människocentrerad, tillförlitlig AI som är känt för etiska och säkra produkter. Detta bygger mer allmänt på de europeiska företagens goda rykte när det gäller att tillhandahålla trygga och säkra produkter av hög kvalitet. Pilotfasen kommer att bidra till att AI-produkterna uppfyller detta löfte.

2.4. Internationella etiska riktlinjer för AI

De internationella diskussionerna om AI-etik har intensifierats efter det att Japans G7-ordförandeskap satte ämnet högt upp på dagordningen 2016. Med tanke på de internationella kopplingarna inom AI-utveckling vad gäller datacirkulation, algoritmisk utveckling och investeringar i forskning **kommer kommissionen att fortsätta sina ansträngningar för att etablera unionens strategi globalt och bygga upp ett samförstånd om människocentrerad artificiell intelligens**[17].

Det arbete som utförts av AI-expertgruppen – mer specifikt förteckningen över krav och samrådsprocessen med berörda parter – ger kommissionen ytterligare värdefull information som grund för de internationella diskussionerna. Europeiska unionen kan ha en ledarroll i utvecklingen av internationella AI-riktlinjer och, om möjligt, en relaterad bedömningsmekanism.

Kommissionen kommer därför att göra följande:

Stärka samarbetet med likasinnade partner genom att

- utforska i vilken utsträckning konvergens kan uppnås med tredjeländers utkast till etiska riktlinjer (t.ex. Japan, Kanada och Singapore) och med utgångspunkt i denna grupp av likasinnade länder förbereda en bredare diskussion med stöd av insatser inom partnerskapsinstrumentet för samarbete med tredjeländer[18], och

---

[17] Unionens höga representant för utrikesfrågor och säkerhetspolitik kommer, med stöd av kommissionen, att fortsätta med samråd i FN, Global Tech-panelen och andra multilaterala sammanhang och framför allt samordna förslag till lösningar på dessa komplexa säkerhetsfrågor.

[18] Europaparlamentets och rådets förordning (EU) nr 234/2014 av den 11 mars 2014 om inrättande av ett partnerskapsinstrument för samarbete med tredjeländer (EUT L 77, 15.3.2014, s. 77). Exempelvis kommer det planerade projektet med en internationell allians för en människocentrerad hållning till artificiell intelligens att underlätta gemensamma initiativ med likasinnade partner för att främja etiska riktlinjer och anta gemensamma principer och operativa slutsatser. Det kommer att göra det möjligt för EU och likasinnade länder att diskutera operativa slutsatser som följer av de etiska riktlinjer om artificiell intelligens som lagts fram av AI-expertgruppen för att nå en gemensam strategi. Dessutom blir det möjligt att övervaka den globala spridningen av AI-teknik. Slutligen planerar man att inom projektet organisera

- utreda hur företag från länder utanför EU och internationella organisationer kan bidra till riktlinjernas ”pilotfas” genom testning och validering.

Fortsätta att spela en aktiv roll i internationella diskussioner och initiativ genom att

- bidra till multilaterala forum som G7 och G20,
- delta i dialoger med länder utanför EU och anordna bilaterala och multilaterala möten för att skapa samförstånd om människocentrerad AI,
- bidra till relevant standardiseringsverksamhet inom internationella standardiseringsorganisationer för att främja denna vision och
- stärka insamlingen och spridningen av insikter om offentlig politik, i samarbete med relevanta internationella organisationer.

3. SLUTSATSER

EU bygger på ett antal grundläggande värden och har byggt upp ett starkt och balanserat regelverk på dessa grunder. Med utgångspunkt i det befintliga regelverket finns det ett behov av etiska riktlinjer för utveckling och användning av artificiell intelligens på grund av att tekniken är så ny och att den innebär särskilda utmaningar. Endast om AI utvecklas och används på ett sätt som respekterar allmänt delade etiska värden kan den betraktas som tillförlitlig.

I strävan efter detta mål välkomnar kommissionen de synpunkter som lagts fram av AI-expertgruppen. Baserat på de krav som är centrala för att AI ska kunna betraktas som tillförlitlig kommer kommissionen nu att inleda en målinriktad pilotfas i syfte att se till att de resulterande etiska riktlinjerna för utveckling och användning av AI kan genomföras i praktiken. Kommissionen kommer också att arbeta för att skapa ett brett samförstånd i samhället om människocentrerad AI, bland annat med alla berörda parter och våra internationella partner.

Den etiska dimensionen av artificiell intelligens är inte någon lyxegenskap eller tilläggsfunktion – den måste vara en integrerad del av utvecklingen av AI. Genom att sträva efter människocentrerad AI som bygger på förtroende skyddar vi respekten för våra centrala samhällsvärden och gör Europa och dess industri till ett utpräglat, ledande varumärke inom banbrytande AI som det går att lita på överallt i världen.

För att säkerställa den etiska utvecklingen av artificiell intelligens i Europa i ett vidare sammanhang strävar kommissionen efter en övergripande strategi som omfattar framför allt följande åtgärder som ska genomföras senast det tredje kvartalet 2019:

- Kommissionen kommer att starta upp ett antal **nätverk av AI-kompetenscentrum** genom Horisont 2020. Den kommer att välja ut upp till fyra nätverk med inriktning på vetenskapliga eller tekniska större utmaningar såsom förklarbarhet och avancerad interaktion mellan människa och maskin, som är viktiga ingredienser för tillförlitlig AI.
- Kommissionen kommer att börja inrätta **nätverk av digitala innovationsknutpunkter**[19] med inriktning på AI inom tillverkning och stordata.

---

offentlig diplomatisk verksamhet i samband med internationella evenemang, t.ex. vid möten inom G7, G20 och Organisationen för ekonomiskt samarbete och utveckling.

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

- Tillsammans med medlemsstaterna och berörda parter kommer kommissionen att inleda förberedande diskussioner för att utveckla och genomföra **en modell för datautbyte och de bästa sätten att använda gemensamma dataområden**, med inriktning framför allt på transport, hälso- och sjukvård samt industriell tillverkning[20].

Dessutom arbetar kommissionen på en rapport om de utmaningar som AI innebär i fråga om säkerhet och ansvar, samt ett vägledande dokument om genomförandet av produktansvarsdirektivet[21]. Samtidigt kommer initiativet för europeiska högpresterande datorsystem (EuroHPC)[22] att utveckla nästa generation av superdatorer, eftersom datorkapacitet är avgörande för att behandla data och lära upp AI, och Europa måste behärska hela den digitala värdekedjan. Det pågående partnerskapet med medlemsstaterna och industrin om mikroelektroniska komponenter och system (Ecsel)[23] och *European Processor Initiative*[24] kommer att bidra till utvecklingen av teknik med lågeffektprocessorer för tillförlitlig och säker högpresterande datateknik och s.k. edge computing.

Liksom arbetet med etiska riktlinjer för artificiell intelligens bygger alla dessa initiativ på ett **nära samarbete mellan alla berörda parter**, såsom medlemsstaterna, näringslivet, samhällsaktörer och enskilda personer. Överlag visar EU:s strategi för artificiell intelligens att ekonomisk konkurrenskraft och samhälleligt förtroende måste utgå från samma grundläggande värden och ömsesidigt förstärka varandra.

[20] De nödvändiga resurserna kommer att mobiliseras från Horisont 2020 (med nästan 1,5 miljarder euro anslagna för AI under perioden 2018–2020) och dess planerade efterföljare Horisont Europa, den digitala delen av Fonden för ett sammanlänkat Europa och framför allt det framtida programmet för ett digitalt Europa. Projekten kommer också att utnyttja resurser från den privata sektorn och medlemsstaternas program.

[21] Se kommissionens meddelande *Artificiell intelligens för Europa*, COM(2018) 237.

[22] https://eurohpc-ju.europa.eu

[23] www.ecsel.eu

[24] www.european-processor-initiative.eu