



## Att skapa förtroende för människocentrerad artificiell intelligens

---

Infrastrukturdepartementet

2019-05-10

### Dokumentbeteckning

KOM (2019) 168

Meddelande om att skapa förtroende för människocentrerad artificiell intelligens

Tidigare faktapromemorior i ärendet:

2017/18:FPM96

Meddelandet om artificiell intelligens för Europa

2018/19:FPM23

Samordnad plan för artificiell intelligens

### Sammanfattning

Kommissionen inrättade 2018 en högnivåexpertgrupp för artificiell intelligens (AI) med företrädare för många olika berörda parter. Högnivågruppen fick bl.a. i uppdrag att ta fram etiska riktlinjer för AI, vilka överlämnades till kommissionen i mars 2019. Bakgrunden är att kommissionen ser ett behov av etiska riktlinjer för AI som bör användas av de som utvecklar, tillhandahåller och använder AI på den inre marknaden. Enligt de riktlinjer som utarbetats av AI-expertgruppen krävs tre komponenter för att ett AI-system ska vara tillförlitligt: det bör vara förenligt med lagstiftningen, det bör uppfylla etiska principer och det bör vara robust. Kommissionen välkomnar AI-expertgruppens arbete och uppmuntrar berörda parter att följa kraven och att testa listan i syfte att bedöma huruvida deras praktiska tillämpning kan skapa rätt förtroendemiljö för en framgångsrik utveckling och användning av artificiell intelligens.

Regeringen välkomnar kommissionens meddelande. Inom ramen för en minskad total budgetram prioriterar regeringen åtgärder för att stödja forskning och innovation, samt de möjligheter digitaliseringen för med sig

och anser att ett gemensamt europeiskt arbete med AI är av stor vikt. Sverige gynnas av att EU utvecklas till en attraktiv och konkurrenskraftig arena där studerande, forskare, företag och offentliga aktörer kan arbeta med utveckling och användning av AI. I enlighet med den nationella inriktningen för AI (N2018/03008/FÖF) anser regeringen att etiska riktlinjer för utveckling och användning av AI är viktigt för att dra nytta av fördelar och samtidigt minimera risker för såväl samhället i stort som för individer och välkomnar därför att kommissionen lyfter etik som ett prioriterat område. Regeringen instämmer i att det är angeläget att främja mångfald, bl.a. avseende kön, inte bara i varje skede i AI-utvecklingen, utan också avseende den data som ligger till grund för AI:s rekommendationer och beslut. Regeringen anser vidare att det är viktigt att det gemensamma arbetet med AI, inklusive utvecklandet av riktlinjer, tar hänsyn till inverkan på medlemsstaternas nationella säkerhet, där medlemsstaternas kompetens på området behöver respekteras.

Regeringen anser att riktlinjer är ett lämpligt instrument för ett område som är under snabb utveckling, i syfte att undvika oproportionerliga regelbördor. Regeringen delar kommissionens uppfattning om vikten av ett konkret och proportionerligt genomförande av riktlinjerna där man utgår från det aktuella sammanhanget vid tillämpningen. Regeringen anser också att det är viktigt att skapa förutsättningar för att testa riktlinjerna på AI-applikationer för att därigenom identifiera eventuella utmaningar och förbättringsområden i syfte att säkerställa att fördelarna med AI kommer till användning. I det arbetet är det viktigt att svenska aktörer har en aktiv roll. Vidare välkomnar regeringen internationellt samarbete kring AI, inklusive etik och hantering av risker med fokus på säkerhet för samhället och individer, då AI drar nytta av gränsöverskridande användning och eftersom EU och Sverige gynnas av internationellt samarbete på området.

## 1 Förslaget

### 1.1 Ärendets bakgrund

I april 2018 presenterade kommissionen en strategi för AI i meddelandet Artificiell intelligens för Europa (KOM(2018) 237) som placerade människor i centrum för utvecklingen av AI – människocentrerad AI. I december 2018 följde kommissionen upp strategin med en samordnad plan om artificiell intelligens (KOM(2018) 795) för att skapa synergier, samla data och öka gemensamma investeringar. Syftet med planen är att främja gränsöverskridande samarbeten och mobilisera berörda aktörer för att öka offentliga och privata investeringar i AI till minst 20 miljarder EUR årligen under de kommande tio åren. Kommissionen har fördubblat sina AI-investeringar inom ramen för forskningsprogrammet Horisont 2020 och planerar att inom ramen för Horisont Europa och programmet för ett digitalt Europa, investera 1 miljard EUR årligen, bl.a. i gemensamma dataområden inom hälsa, transport och tillverkning och stora experimentanläggningar som

smarta sjukhus och infrastrukturer för automatiserade fordon och en strategisk forskningsagenda. I syfte att genomföra en sådan gemensam strategisk forsknings-, innovations- och genomförandeagenda har kommissionen intensifierat dialogen med berörda parter från näringsliv, forskningsinstitut och myndigheter.

Både den europeiska AI-strategin och den samordnade planen för AI framhåller att tillit är en förutsättning för ett människocentrerat förhållningssätt till AI. Kommissionen menar att Europas etiska förhållningssätt till AI stärker medborgares tillit till den digitala utvecklingen och skapar en konkurrensfördel för europeiska AI-företag. Syftet med det nu aktuella meddelandet är att komplettera AI-strategin och den samordnade planen bl.a. genom att lansera en övergripande pilotfas som inkluderar många olika aktörer för att testa det praktiska genomförandet av etiska riktlinjer för utveckling och användning av AI.

Kommissionen understryker att AI inte är ett självändamål i sig utan ett verktyg som ska hjälpa människor, med det yttersta målet att öka människors välbefinnande. Utöver EU:s stadga om de grundläggande rättigheterna och de värden som unionen bygger på – respekt för människans värdighet, frihet, demokrati, jämlikhet, rättsstaten och respekt för de mänskliga rättigheterna, så har EU ett starkt regelverk som enligt kommissionen kommer att utgöra en global standard för människocentrerad AI. Det inkluderar EU:s dataskyddsförordning, förordningen om det fria flödet av uppgifter som inte är personuppgifter och cybersäkerhetsakten.

Användningen av AI innebär dock nya utmaningar eftersom det möjliggör för maskiner att "lära sig" och att fatta och genomföra beslut utan mänsklig inblandning. Beslut som fattas av algoritmer kan vara baserade på data som är ofullständiga och därför inte tillförlitliga, de kan vara manipulerade genom cyberangrepp och de kan vara snedvridna eller på andra sätt felaktiga. Om tekniken används utan reflektion kan det innebära problematiska resultat och leda till tveksamhet hos allmänheten vad gäller deras acceptans för AI. Kommissionen konstaterar därför att AI-tekniken behöver utvecklas med människor i centrum för att därigenom förtjäna allmänhetens förtroende. Mångfald i fråga om kön, ras eller etniskt ursprung, religion eller övertygelse, funktionsvariation och ålder bör garanteras i varje skede av AI-utvecklingen. Syftet bör vara att förbättra människors förmågor, inte att ersätta dem, och de bör även vara tillgängliga för personer med funktionsnedsättning. Enligt kommissionen innebär det här att AI-applikationer inte bara ska vara förenliga med lagstiftningen utan också följa etiska principer samt se till att oavsiktliga skador undviks i praktiken.

Kommissionen har därför sett ett behov av etiska riktlinjer som bygger på det gällande rättsliga ramverket och som bör användas av de som utvecklar, tillhandahåller och använder AI på den inre marknaden. Kommissionen inrättade följaktligen en högnivåexpertgrupp för AI med företrädare för många olika berörda parter som bl.a. fått i uppdrag att ta fram etiska riktlinjer för AI, vilka överlämnades till kommissionen i mars 2019.

Enligt de riktlinjer som utarbetats av AI-expertgruppen, och som hänvisas till i kommissionens meddelande, krävs tre komponenter för att AI ska vara tillförlitlig: den bör vara förenlig med lagstiftningen, den bör uppfylla etiska principer och den bör vara robust. Utifrån dessa tre komponenter, anges i riktlinjerna sju centrala krav som AI-tillämpningar bör uppfylla för att betraktas som tillförlitliga:

- mänskligt agentskap och mänsklig tillsyn
- teknisk robusthet och säkerhet
- integritet och dataförvaltning
- transparens
- mångfald, icke-diskriminering och rättvisa
- samhällets och miljöns välbefinnande
- ansvarsskyldighet.

I meddelandet framgår att även om kraven är tänkta att gälla alla AI-system inom olika miljöer och branscher bör hänsyn tas till det specifika sammanhang där de används för att säkra ett konkret och proportionellt genomförande, med utgångspunkt i applikationens inverkan på sin omgivning.

Riktlinjerna är inte bindande och skapar i sig inga nya rättsliga skyldigheter. Dock återspeglar många befintliga bestämmelser i gällande unionsrätt ett eller flera av dessa centrala krav, såsom regler om säkerhet, skydd av personuppgifter, integritet eller miljöskydd.

Kommissionen konstaterar att ett viktigt steg på vägen mot riktlinjer för etisk AI är att nå samförstånd kring kraven för AI-system. En del i detta är att testa riktlinjerna. I ett nästa steg kommer kommissionen därför att se till att riktlinjerna kan testas och genomföras i praktiken. Kommissionen kommer att genomföra detta dels genom en riktad pilotfas som är utformad för att få strukturerad återkoppling från berörda parter, dels genom att fortsätta samråd med berörda parter samt genom att genomföra medvetandehöjande åtgärder i medlemsstaterna och bland olika grupper av berörda parter, bl.a. inom industri- och tjänstesektorena.

I början av 2020 kommer AI-expertgruppen att se över och uppdatera riktlinjerna. På grundval av översynen och de erfarenheter som gjorts kommer kommissionen att utvärdera resultatet och föreslå nästa steg.

Som en konsekvens av de internationella kopplingarna inom AI-utveckling vad gäller exempelvis data-cirkulation, algoritmisk utveckling och investeringar i forskning så kommer kommissionen att fortsätta arbetet med att etablera unionens strategi globalt och bygga ett samförstånd om människocentrerad AI. Kommissionen avser därför att 1) stärka samarbetet

med likasinnade parter genom att undersöka i vilken utsträckning överensstämmelse kan uppnås med tredjeländers utkast till etiska riktlinjer (t.ex. Japan, Kanada och Singapore), 2) utreda hur företag från länder utanför EU och internationella organisationer kan bidra i riktlinjernas pilotfas genom testning och validering. Kommissionen kommer därtill att fortsätta att spela en aktiv roll i internationella diskussioner och initiativ.

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

- kommissionen kommer att starta ett antal nätverk av AI-kompetenscentrum genom Horisont 2020. Upp till fyra nätverk kommer att väljas ut med inriktning på vetenskapliga eller tekniska större utmaningar såsom förklarbarhet och avancerad interaktion mellan människa och maskin, som är viktiga komponenter för tillförlitlig AI,
- kommissionen kommer att påbörja inrättandet av ett nätverk av digitala innovationsknutpunkter med inriktning på AI inom tillverkning och stordata,
- tillsammans med medlemsstaterna och berörda parter kommer kommissionen att inleda förberedande diskussioner för att utveckla och genomföra en modell för datautbyte och identifiera de bästa sätten att använda gemensamma dataområden, med inriktning framför allt på transport, hälso- och sjukvård samt industriell tillverkning.

Kommissionen arbetar därtill vidare med en rapport om de utmaningar AI innebär i fråga om regelverken för säkerhet och ansvar, samt ett vägledande dokument till produktansvarsdirektivet.

1.3 Gällande svenska regler och förslagets effekt på dessa Riktlinjerna är inte bindande och skapar i sig inga nya rättsliga skyldigheter.

1.4 Budgetära konsekvenser / Konsekvensanalys

Förslaget bedöms inte ha budgetära konsekvenser.

## 2 Ståndpunkter

### 2.1 Preliminär svensk ståndpunkt

Regeringen välkomnar kommissionens meddelande. Inom ramen för en minskad total budgetram prioriterar regeringen åtgärder för att stödja

forskning och innovation, samt de möjligheter digitaliseringen för med sig och anser att ett gemensamt europeiskt arbete med AI är av stor vikt. Sverige gynnas av att EU utvecklas till en attraktiv och konkurrenskraftig arena där studerande, forskare, företag och offentliga aktörer kan arbeta med utveckling och användning av AI. Stora värden kan skapas för samhället som helhet när offentlig och privat sektor använder AI-lösningar på en rad områden.

I enlighet med den nationella inriktningen för AI (N2018/03008/FÖF) anser regeringen att etiska riktlinjer för utveckling och användning av AI är viktigt för att dra nytta av fördelar och samtidigt minimera risker för såväl samhället i stort som för individer och välkomnar därför att kommissionen lyfter etik som ett prioriterat område. Regeringen instämmer i att det är angeläget att främja mångfald, bl.a. avseende kön, inte bara i varje skede i AI-utvecklingen, utan också avseende den data som ligger till grund för AI:s rekommendationer och beslut. Regeringen anser vidare att det är viktigt att det gemensamma arbetet med AI, inklusive utvecklandet av riktlinjer, tar hänsyn till inverkan på medlemsstaternas nationella säkerhet, där medlemsstaternas kompetens på området behöver respekteras.

Regeringen anser att riktlinjer är ett lämpligt instrument för ett område som är under snabb utveckling, i syfte att undvika oproportionerliga regelbördor. Regeringen delar kommissionens uppfattning om vikten av ett konkret och proportionerligt genomförande av riktlinjerna där man utgår från det aktuella sammanhanget i tillämpningen. Det är viktigt att det finns mekanismer som säkerställer såväl möjligheten till granskning som skydd av affärsmodeller och immateriella rättigheter. Regeringen anser också att det är viktigt att skapa förutsättningar för att testa riktlinjerna på AI-applikationer för att därigenom identifiera eventuella utmaningar och förbättringsområden för att säkerställa att fördelarna med AI kommer till användning. I det arbetet är det viktigt att svenska aktörer har en aktiv roll. Vidare välkomnar regeringen internationellt samarbete kring AI, inklusive etik, och hantering av risker med fokus på säkerhet för samhället och individer, då AI drar nytta av gränsöverskridande användning och eftersom EU och Sverige gynnas av internationellt samarbete på området.

## 2.2 Medlemsstaternas ståndpunkter

Den 11 februari 2019 antog rådet slutsatser där man välkomnar kommissionens meddelande av den 7 december 2018 om en samordnad plan om artificiell intelligens och efterlyser en starkare utveckling, ett kraftfullare införande och en ökad användning av tillämpningar för artificiell intelligens i alla ekonomiska sektorer, i syfte att göra Europa till en global ledare inom AI. Vidare betonar slutsatserna vikten av att man säkerställer att EU-medborgarnas rättigheter respekteras fullt ut, genom att införa etiska riktlinjer för utvecklingen och användningen av artificiell intelligens i EU och på global nivå, och på så sätt gör etik inom AI till en konkurrensfördel för den europeiska industrin. Medlemsstaterna noterar i rådslutsatserna att den oberoende högnivåexpertgruppen för AI kommer att offentliggöra etiska

riktlinjer och ser fram emot att driva arbetet vidare. Medlemsstaterna stöder vidare kommissionens arbete med att etablera EU:s etiska strategi globalt och verkar för att den främjas inom internationella initiativ för AI, däribland OECD:s och den internationella panelen för artificiell intelligens initiativ.

### 2.3 Institutionernas ståndpunkter

Institutionernas ståndpunkter är inte kända.

### 2.4 Remissinstansernas ståndpunkter

Meddelandet har inte sänts på remiss.

## 3 Förslagets förutsättningar

### 3.1 Rättslig grund och beslutsförfarande

Inte aktuellt då meddelandet inte innehåller några förslag till lagstiftningsakter.

### 3.2 Subsidiaritets- och proportionalitetsprincipen

Inte aktuellt. Meddelandet innehåller inga förslag till lagstiftningsakter.

## 4 Övrigt

### 4.1 Fortsatt behandling av ärendet

Rådets ordförandeskap har inte ännu aviserat någon vidare behandling av meddelandet.

### 4.2 Fackuttryck/termer

Artificiell intelligens = Artificiell intelligens (AI) har ingen entydig definition eller allmänt vedertagen avgränsning, utan det finns flera definitioner. I allmänhet avses emellertid intelligens som uppvisas av maskiner, eller förmågan hos en maskin att efterlikna intelligent mänskligt beteende. (Vinnovas rapport "Artificiell intelligens i svenskt näringsliv och samhälle" (2017/07836/FÖF)).

Mänskligt agentskap = AI-riktlinjerna beskriver mänskligt agentskap som innebärande att användare kan fatta informerade, autonoma beslut angående AI-system. De bör få den kunskap och de verktyg som krävs för att förstå och interagera med AI-system för att, när det är möjligt, kunna utvärdera eller ifrågasätta systemet. AI-system ska stödja individer i att fatta bättre och mer underbyggda beslut i enlighet med sina mål. Den övergripande principen om användarens autonomi måste vara central i systemets funktioner.