EUROPEISKA
KOMMISSIONEN

Bryssel den 19.2.2020
COM(2020) 65 final

VITBOK

Om artificiell intelligens - en EU-strategi för spetskompetens och förtroende

Vitbok om artificiell intelligens
En EU-strategi för spetskompetens och förtroende

Artificiell intelligens är en teknik under snabb utveckling. Den kommer att förändra våra liv genom att förbättra hälso- och sjukvården (till exempel genom exaktare diagnoser för att bättre förebygga sjukdomar), effektivisera jordbruket, begränsa klimatförändringarna och förbättra anpassningen till dem, effektivisera produktionssystemen med hjälp av prediktivt underhåll, göra EU säkrare för medborgarna, och på många andra sätt som vi i dag knappt kan föreställa oss. Men artificiell intelligens (AI) medför också ett antal risker, till exempel otydliga beslutsvägar, könsdiskriminering eller andra former av diskriminering, intrång i privatlivet eller användning för brottsliga ändamål.

I den rådande hårda globala konkurrensen behövs en solid EU-strategi som bygger på den europeiska strategin för AI, som lades fram i april 2018[1]. För att utnyttja möjligheterna med AI och hantera utmaningarna måste EU agera samfällt och definiera ett eget sätt att främja utvecklingen och spridningen av AI som grundar sig på europeiska värderingar.

Kommissionen är fast besluten att möjliggöra vetenskapliga genombrott, bevara EU:s tekniska ledarställning och se till att nya tekniska lösningar är till nytta för alla européer, förbättrar deras liv och samtidigt respekterar deras rättigheter.

I sina politiska riktlinjer[2] tillkännagav kommissionens ordförande Ursula von der Leyen en samordnad europeisk strategi för de mänskliga och etiska konsekvenserna av artificiell intelligens, och en reflektion om hur stordata bäst kan användas för innovationer.

Kommissionen stöder därför en lösning som bygger på lagstiftning och som är investeringsorienterad, med den dubbla målsättningen att främja användningen av AI och ta itu med riskerna förknippade med viss användning av den nya tekniken. Syftet med den här vitboken är att beskriva olika politiska alternativ för att nå de här målen. Utveckling och användning av AI för militära ändamål behandlas inte. Kommissionen uppmanar medlemsstaterna, andra europeiska institutioner och alla aktörer, inbegripet industrin, arbetsmarknadens parter, organisationer i det civila samhället, forskare, allmänheten och andra berörda parter, att inkomma med reaktioner på nedanstående alternativ och bidra till kommissionens framtida beslutsfattande på det här området.

1. INLEDNING

Digitaltekniken har blivit ett allt vanligare inslag i alla delar av tillvaron. Därför bör människor kunna lita på den. Tillförlitligheten är också en förutsättning för dess spridning. Det här är ett tillfälle för EU med tanke på den vikt som EU fäster vid värderingar och rättsstatsprincipen, och med tanke på EU:s dokumenterade förmåga att ta fram säkra, tillförlitliga och sofistikerade produkter och tjänster inom allt från flygteknik till energi, fordonsindustri och medicinsk utrustning.

EU:s nuvarande och framtida hållbara ekonomiska tillväxt och sociala välfärd bygger i allt högre grad på värde som skapas av data. AI är en av dataekonomins viktigaste tillämpningar. Dagens data rör oftast konsumenter och lagras och behandlas i centrala molnbaserade infrastrukturer. Morgondagens mycket mer omfattande data kommer i stället att komma från industrin, näringslivet och den offentliga sektorn, och kommer att lagras i en rad olika system, särskilt datorutrustning som fungerar i nätets

[1] Artificiell intelligens för Europa, COM(2018) 237 final.
[2] https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_sv.pdf

periferi. Det här skapar nya möjligheter för EU, som har en stark ställning inom digitaliserade industri- och företag-till-företag-tillämpningar men en relativt svag ställning inom konsumentplattformar.

Enkelt uttryckt är AI en samling tekniker som kombinerar data, algoritmer och datorkapacitet. Framsteg inom databehandling och den ökande tillgången till data är därför viktiga drivkrafter bakom AI:s nuvarande expansion. **EU kan bli världsledande inom innovation i dataekonomin och dess tillämpningar** genom att kombinera sin tekniska och industriella styrka med en digital infrastruktur av hög kvalitet och ett regelverk baserat på grundläggande europeiska värden, såsom anges i EU:s datastrategi[3]. På grundval av detta kan unionen utveckla ett AI-ekosystem som gör så att fördelarna med de tekniska lösningarna kommer till nytta för hela det europeiska samhället och den europeiska ekonomin:

- för **medborgarna**, som kan dra nytta av de nya fördelarna med till exempel förbättrad hälso- och sjukvård, färre trasiga hushållsmaskiner, säkrare och renare transportsystem och bättre offentliga tjänster,
- för **näringslivets** utveckling, till exempel genom en ny generation produkter och tjänster på områden där EU är särskilt starkt (maskiner, transporter, cybersäkerhet, jordbruk, den gröna och cirkulära ekonomin, hälso- och sjukvård samt sektorer med högt mervärde såsom mode och turism), och
- för tjänster av **allmänt intresse**, till exempel genom att minska kostnaderna för att tillhandahålla olika tjänster (transport, utbildning, energi och avfallshantering), genom att öka produkternas hållbarhet[4] och genom att förse de brottsbekämpande myndigheterna med lämpliga verktyg för att garantera medborgarnas säkerhet[5], tillsammans med lämpliga skyddsåtgärder för att respektera deras fri- och rättigheter.

Med tanke på den stora inverkan som AI kan få på vårt samhälle och behovet av att skapa förtroende är det viktigt att den europeiska AI:n bygger på europeiska värden och grundläggande rättigheter såsom mänsklig värdighet och integritet.

Vidare bör man inte endast beakta hur AI-systemen påverkar enskilda, utan även hur de påverkar samhället i stort. Användningen av AI-system kan spela en viktig roll för att uppnå målen för hållbar utveckling och för att stödja den demokratiska processen och de sociala rättigheterna. Genom förslagen i den europeiska gröna given[6] nyligen visar EU vägen när det gäller att ta itu med klimat- och miljörelaterade utmaningar. Den digitala tekniken, till exempel AI, är en viktig förutsättning för att uppnå målen i den gröna given. Med tanke på AI:s ökande betydelse måste AI-systemens miljöpåverkan beaktas under hela deras livscykel och i hela leveranskedjan, till exempel när det gäller resursanvändningen för att träna algoritmer och lagra data.

En gemensam europeisk strategi för AI är nödvändig för att uppnå tillräcklig omfattning och undvika en fragmentering av den inre marknaden. Nationella initiativ skulle riskera att äventyra rättssäkerheten, försvaga medborgarnas förtroende och förhindra framväxten av en dynamisk europeisk industri.

[3] COM(2020) 66 final.
[4] AI och digitalisering i allmänhet är viktiga förutsättningar för att målen i den europeiska gröna given ska uppnås. Emellertid uppskattas IKT-sektorns nuvarande miljöavtryck utgöra mer än 2 % av de globala utsläppen. I EU:s digitala strategi, som åtföljer den här vitboken, föreslås gröna omställningsåtgärder för digital teknik.
[5] AI-verktyg kan erbjuda en möjlighet att bättre skydda EU-medborgare från brottslighet och terroristdåd. Sådana verktyg skulle till exempel kunna användas för att identifiera terroristpropaganda på nätet, upptäcka misstänkta transaktioner vid försäljning av farliga produkter, identifiera farliga gömda föremål eller olagliga ämnen eller produkter, bistå medborgare i nödsituationer och vägleda insatspersonal.
[6] COM(2019) 640 final.

I den här vitboken anges olika politiska alternativ för att möjliggöra en tillförlitlig och säker utveckling av AI i Europa, med full respekt för EU-medborgarnas värderingar och rättigheter. De viktigaste byggstenarna i den här vitboken är följande:

- Den politiska ramen med åtgärder för att samordna insatserna på europeisk, nationell och regional nivå. Syftet med ramen är att, genom partnerskap mellan den privata och den offentliga sektorn, mobilisera resurser för att uppnå ett **”ekosystem av spetskompetens”** längs hela värdekedjan, med början i forskning och innovation, och att skapa de rätta incitamenten för att påskynda införandet av lösningar som baseras på AI, även i små och medelstora företag.
- De viktigaste delarna i ett framtida regelverk för AI i Europa som kommer att skapa ett unikt **”ekosystem av förtroende”.** För att åstadkomma detta måste regelverket säkerställa att EU:s regler följs, till exempel reglerna om skydd av de grundläggande rättigheterna och konsumenträttigheterna, i synnerhet när det gäller AI-system som används inom EU och som utgör en hög risk[7]. Att skapa ett ekosystem av förtroende är ett politiskt mål i sig, och det bör ge medborgarna förtroende att använda AI-tillämpningar, och skapa rättslig säkerhet för företag och offentliga organisationer att driva innovation med hjälp av AI. Kommissionen stöder starkt en människocentrerad strategi som bygger på meddelandet om att skapa förtroende för människocentrerad artificiell intelligens[8] och kommer också att beakta de bidrag som erhållits under pilotfasen för de etiska riktlinjer som utarbetats av expertgruppen på hög nivå för AI-frågor.

Syftet med den europeiska strategin för data, som åtföljer den här vitboken, är att göra det möjligt för EU att bli den mest attraktiva, säkra och dynamiska data-agila ekonomin i världen och därigenom ge EU möjlighet att med hjälp av data fatta bättre beslut och förbättra medborgarnas liv. Strategin innehåller ett antal politiska åtgärder, bland annat mobilisering av privata och offentliga investeringar, som behövs för att uppnå det här målet. De konsekvenser som AI, sakernas internet och annan digital teknik har för säkerhets- och ansvarslagstiftningen analyseras slutligen i den kommissionsrapport som åtföljer den här vitboken.

2. UTNYTTJA STYRKAN PÅ INDUSTRINS OCH FÖRETAGENS MARKNADER

EU har goda förutsättningar att dra nytta av potentialen hos AI, inte endast som användare, utan även som upphovsman och framställare av AI-teknik. EU har utmärkta forskningscentrum, innovativa nyetablerade företag, en världsledande ställning inom robotteknik och konkurrenskraftiga tillverknings- och tjänstesektorer, från bilindustri till hälso- och sjukvård, energi, finansiella tjänster och jordbruk. EU har utvecklat en stark datorinfrastruktur (till exempel högpresterande datorer), vilket är nödvändigt för att AI ska fungera. EU har också stora volymer offentliga och industriella data, vars potential för närvarande underutnyttjas. Det finns en erkänd industriell styrka inom säkra digitala system med låg energiförbrukning, vilket är avgörande för vidareutvecklingen av AI.

Om EU:s kapacitet att investera i nästa generation av teknik och infrastruktur och i digital kompetens såsom datakunskap utnyttjas, kommer det att öka EU:s tekniska oberoende i fråga om viktig möjliggörande teknik och infrastruktur för dataekonomin. Infrastrukturen bör stödja skapandet av

[7] Även om ytterligare åtgärder kan behöva vidtas för att förebygga och motverka användning av AI för brottsliga ändamål, faller detta utanför den här vitbokens tillämpningsområde.
[8] COM(2019) 168.

europeiska datapooler som möjliggör tillförlitlig AI, till exempel AI som är baserad på europeiska värden och regler.

EU bör utnyttja sina starka sidor för att utvidga sin position i ekosystemen och längs värdekedjan, från vissa sektorer för maskinvarutillverkning via programvara hela vägen till tjänster. Detta sker redan i viss mån. EU producerar mer än en fjärdedel av alla robotar för industriella och yrkesinriktade tillämpningar (till exempel för precisionsjordbruk, säkerhet, hälsa och logistik) och spelar en viktig roll i utveckling och användning av programvarutillämpningar för företag och organisationer (företag-till-företag-tillämpningar såsom programvara för resursplanering, design och teknik) samt tillämpningar för att stödja e-förvaltning och det ”intelligenta företaget”.

EU är ledande när det gäller att sprida användningen av AI inom tillverkning. Mer än hälften av de ledande tillverkarna tillämpar åtminstone ett moment av artificiell intelligens i tillverkningsprocessen[9].

En anledning till Europas starka ställning när det gäller forskning är EU:s finansieringsprogram som har visat sig vara avgörande för att samla insatserna, undvika dubbelarbete och dra nytta av offentliga och privata investeringar i medlemsstaterna. Under de senaste tre åren har EU:s finansiering av forskning och innovation inom AI ökat till 1,5 miljarder euro, det vill säga en ökning med 70 % jämfört med föregående period.

Investeringarna i forskning och innovation i EU är dock fortfarande en bråkdel av de offentliga och privata investeringarna i andra regioner i världen. Omkring 3,2 miljarder euro investerades i AI i Europa 2016, jämfört med omkring 12,1 miljarder euro i Nordamerika och 6,5 miljarder euro i Asien[10]. Som svar på detta måste EU öka sina investeringsnivåer avsevärt. Den samordnade AI-plan[11] som utarbetats tillsammans med medlemsstaterna har visat sig vara en god utgångspunkt när det gäller att bygga upp ett närmare samarbete om AI i Europa och skapa synergieffekter för att maximera investeringarna i AI-värdekedjan.

3. UTNYTTJA FRAMTIDENS MÖJLIGHETER: NÄSTA VÅG AV DATA

Även om Europa för närvarande har en svagare ställning när det gäller konsumenttillämpningar och onlineplattformar, vilket resulterar i en konkurrensnackdel när det gäller tillgång till data, är stora förändringar på väg i fråga om värdet och återanvändandet av gemensamma data för olika sektorer. Den mängd data som produceras runtom i världen ökar snabbt, från 33 zettabyte 2018 till förväntade 175 zettabyte 2025[12]. Varje ny våg av data ger EU möjlighet att positionera sig inom den data-agila ekonomin och bli världsledande på det här området. Hur lagring och behandling av data sker kommer dessutom att förändras dramatiskt under de kommande fem åren. Fyra femtedelar av all databehandling och dataanalys i molnet i dag görs i datacentraler och centraliserade datoranläggningar, medan en femtedel görs i smarta uppkopplade föremål, såsom bilar, hushållsapparater och tillverkningsrobotar, samt i datoranläggningar nära användaren (s.k. *edge computing*). Fram till 2025 kommer de här andelarna sannolikt att ha ändrats markant[13].

EU är världsledande när det gäller lågeffektelektronik, som är avgörande för nästa generation specialiserade processorer för AI. Den här marknaden domineras för närvarande av aktörer utanför EU. Det här skulle kunna förändras med hjälp av initiativ som till exempel Europeiska processorinitiativet, som är inriktat på utveckling av datasystem med låg energiförbrukning för både perifer och nästa

[9] Följt av Japan (30 %) och USA (28 %). Källa: Capgemini (2019).
[10] *10 imperatives for Europe in the age of AI and automation*, McKinsey (2017).
[11] COM(2018) 795.
[12] IDC (2019).
[13] Gartner (2017).

generationens högpresterande databehandling, och arbetet i det gemensamma företaget *Key Digital Technology Joint Undertaking*, som föreslås starta 2021. Europa är också ledande inom neuromorfiska lösningar [14] som är idealiska för att automatisera industriella processer (”Industri 4.0”) och transportslag. De kan förbättra energieffektiviteten flerfaldigt.

Den senaste tidens framsteg inom kvantdatortekniken kommer att generera en exponentiell ökning av databehandlingskapaciteten[15]. EU kan ta ledningen inom den här tekniken tack vare sin akademiska styrka inom kvantdatorteknik, liksom den europeiska industrins starka ställning inom kvantsimulatorer och programmeringsmiljöer för kvantdatorteknik. EU-initiativ som syftar till att öka tillgången till testnings- och experimentanläggningar för kvantteknik kommer att bidra till tillämpningen av de här nya kvantlösningarna i ett antal industrisektorer och akademiska sektorer.

Samtidigt kommer EU fortsätta att leda utvecklingen när det gäller de algoritmiska grunderna för AI, grundat på sin egen vetenskapliga spetskompetens. Det finns ett behov av att bygga broar mellan ämnesområden där arbetet för närvarande sker separat, till exempel maskininlärning och djupinlärning (som kännetecknas av begränsad tolkningsbarhet, behovet av stora dataset för att träna modeller och för att lära genom korrelationer) och symboliska tillvägagångssätt (där regler skapas genom mänskligt ingripande). Att kombinera ett symboliskt resonemang med djupinlärande neuronnät kan hjälpa oss att förbättra förklarbarheten hos AI-resultat.

4. ETT EKOSYSTEM AV SPETSKOMPETENS

För att bygga upp ett ekosystem av spetskompetens som kan stödja utvecklingen och användningen av AI inom hela EU:s ekonomi och offentliga förvaltning finns det ett behov av att intensifiera åtgärderna på flera nivåer.

A. SAMARBETE MED MEDLEMSSTATERNA

Som utlovat i AI-strategin som antogs i april 2018[16] lade kommissionen i december 2018 fram en samordnad plan, som utarbetats tillsammans med medlemsländerna, för att främja utvecklingen och användningen av AI i Europa[17].

I den här planen föreslås cirka 70 gemensamma åtgärder för närmare och effektivare samarbete mellan medlemsstaterna och kommissionen på viktiga områden såsom forskning, investeringar, marknadspenetration, kompetens och begåvning, data och internationellt samarbete. Planen beräknas löpa fram till 2027, med regelbunden övervakning och översyn.

Syftet är att maximera effekten av investeringar i forskning, innovation och spridning, bedöma nationella AI-strategier och bygga vidare på och utvidga den samordnade planen om AI med medlemsstaterna.

- *Åtgärd 1: Kommissionen kommer, med beaktande av resultaten av det offentliga samrådet om vitboken, att föreslå medlemsstaterna en översyn av den samordnade planen som ska antas före utgången av 2020.*

[14] Med neuromorfiska lösningar menas mycket storskaliga system av integrerade kretsar som efterliknar de neurobiologiska arkitekturer som finns i nervsystemet.

[15] Kvantdatorer kommer att ha kapacitet att på någon sekund bearbeta mångdubbelt större dataset än dagens mest högpresterande datorer, vilket gör det möjligt att utveckla nya AI-tillämpningar i olika sektorer.

[16] Artificiell intelligens för Europa, COM(2018) 237.

[17] Samordnad plan om artificiell intelligens, COM(2018) 795.

EU-finansiering av AI bör locka till sig och samla investeringarna i områden där de åtgärder som krävs går längre än vad som kan uppnås av varje medlemsstat separat. Målet är att locka mer än 20 miljarder euro[18] i totala AI-investeringar i EU per år under det kommande årtiondet. För att stimulera privata och offentliga investeringar kommer EU att tillhandahålla resurser från programmet för ett digitalt Europa, Horisont Europa och de europeiska struktur- och investeringsfonderna för att tillgodose behoven i mindre utvecklade regioner samt landsbygdsområden.

Den samordnade planen skulle också kunna behandla samhälleligt och miljömässigt välbefinnande som en huvudprincip för AI. AI-systemen kommer sannolikt att bidra till att hantera de mest akuta problemen, bland annat klimatförändringar och miljöförstöring. Det är också viktigt att detta sker på ett miljövänligt sätt. AI kan och bör på egen hand kritiskt granska resursanvändning och energiförbrukning och tränas att göra val som är positiva för miljön. Tillsammans med medlemsstaterna kommer kommissionen att överväga olika alternativ för att uppmuntra och främja AI-lösningar som gör detta.

B. FOKUSERADE INSATSER INOM FORSKNING OCH INNOVATION

Europa har inte råd att upprätthålla den nuvarande fragmenteringen med individuella kompetenscentrum som vart och ett inte är tillräckligt storskaligt för att kunna konkurrera med de ledande instituten på global nivå. Det är absolut nödvändigt att skapa fler synergier och nätverk mellan de många europeiska AI-forskningscentrumen och samordna deras insatser för att förbättra spetskompetens, behålla och locka till sig de bästa forskarna och utveckla den bästa tekniken. Europa behöver ett centrum för forskning, innovation och spetskompetens som en ledstjärna som kan samordna de här insatserna och utgöra en världsreferens när det gäller spetskompetens inom AI och som kan locka till sig investeringar och de bästa talangerna på området.

Centrumen och nätverken bör koncentreras till sektorer där EU har potential att bli världsledande, såsom industri, hälso- och sjukvård, transporter, finansiering, värdekedjor för jordbruksbaserade livsmedel, energi/miljö, skogsbruk, jordobservation och rymden. På alla de här områdena pågår kapplöpningen om globalt ledarskap, och EU kan erbjuda stor potential, kunskap och expertis[19]. Lika viktigt är att skapa testnings- och experimentanläggningar för att stödja utvecklingen och den efterföljande spridningen av nya AI-tillämpningar.

- *Åtgärd 2: Kommissionen kommer, eventuellt med hjälp av ett nytt rättsligt instrument, att underlätta skapandet av spetskompetens- och testcenter som kan förena europeiska, nationella och privata investeringar. Kommissionen har föreslagit ett ambitiöst och öronmärkt belopp för att inom EU stödja testcenter som utgör global referens inom ramen för programmet för ett digitalt Europa, vid behov kompletterat med forsknings- och innovationsåtgärder inom Horisont Europa som en del av den fleråriga budgetramen för 2021–2027.*

C. KOMPETENS

Den europeiska strategin för AI måste stödjas av en stark inriktning på att åtgärda kompetensbrister.[20] Kommissionen kommer snart att lägga fram en förstärkt kompetensagenda, vilken har till syfte att se

[18] COM(2018) 237.

[19] Den framtida Europeiska försvarsfonden och det permanenta strukturerade samarbetet (Pesco) kommer också att ge möjligheter till forskning och utveckling inom AI. De här projekten bör synkroniseras med EU:s bredare civila program för AI.

[20] https://ec.europa.eu/jrc/en/publication/academic-offer-and-demand-advanced-profiles-eu

till att alla i EU kan dra nytta av den gröna respektive digitala omställningen av EU:s ekonomi. Initiativen skulle också kunna omfatta stöd till regleringsmyndigheter för olika sektorer för att förbättra deras AI-kompetens och därigenom på ett effektivt och ändamålsenligt sätt genomföra relevanta regler. Den uppdaterade handlingsplanen för digital utbildning kommer att bidra till ett bättre utnyttjande av data och AI-baserad teknik, till exempel lärandeanalys och prediktiv analys, i syfte att förbättra utbildningssystemen och anpassa dem till den digitala tidsåldern. Handlingsplanen kommer också att öka medvetenheten om AI på alla utbildningsnivåer för att förbereda medborgarna på att beslutsunderlag i allt högre grad kommer att påverkas av AI.

Att utveckla den kompetens som krävs för att arbeta inom AI och höja arbetskraftens kompetens så att den passar den AI-ledda omvandlingen kommer att prioriteras i den reviderade samordnade AI-planen som ska utarbetas tillsammans med medlemsstaterna. Detta skulle kunna inbegripa en omvandling av bedömningslistan i de etiska riktlinjerna till en vägledande ”läroplan” för utvecklare av AI som kommer att göras tillgänglig som en resurs för utbildningsinstitutioner. Särskilda ansträngningar bör göras för att öka antalet kvinnor som utbildas och anställs på det här området.

Dessutom skulle ett centrum för forskning och innovation för AI som en ledstjärna i Europa locka begåvning från hela världen på grund av de möjligheter det skulle kunna erbjuda. Det skulle också utveckla och sprida spetskompetens som slår rot och växer i hela Europa.

- *Åtgärd 3: Bilda och stödja, med hjälp av avancerad kompetens-pelaren i programmet för ett digitalt Europa, nätverk av ledande universitet och högskolor för att locka till sig de bästa professorerna och forskarna och erbjuda världsledande masterprogram inom AI.*

Utöver kompetenshöjningen påverkas arbetstagare och arbetsgivare direkt även av utformningen och användningen av AI-system på arbetsplatsen. Deltagande av arbetsmarknadens parter kommer att vara en avgörande faktor för att säkerställa en människocentrerad syn på AI på arbetsplatsen.

D. INRIKTNING PÅ SMÅ OCH MEDELSTORA FÖRETAG

Det kommer också att vara viktigt att se till att små och medelstora företag kan få tillgång till och använda AI. Därför bör de digitala innovationsknutpunkterna[21] och plattformen för efterfrågestyrd AI[22] stärkas ytterligare och främja samarbetet mellan små och medelstora företag. Programmet för ett digitalt Europa kommer att vara avgörande för att uppnå detta. Även om alla digitala innovationsknutpunkter bör ge stöd till de små och medelstora företagen så att de förstår och tar sig an AI, kommer det att vara viktigt att minst en innovationsknutpunkt per medlemsstat har en hög grad av specialisering inom AI.

Små och medelstora företag och nystartade företag kommer att behöva tillgång till finansiering för anpassning av sina processer eller för innovation med hjälp av AI. Med utgångspunkt i den kommande pilotinvesteringsfonden på 100 miljoner euro i AI och blockkedjeteknik planerar kommissionen att ytterligare öka tillgången till finansiering för AI inom ramen för InvestEU[23]. AI nämns uttryckligen bland de stödberättigade områdena för användning av InvestEU-garantin.

- *Åtgärd 4: Kommissionen kommer att samarbeta med medlemsstaterna för att se till att minst en digital innovationsknutpunkt per medlemsstat har en hög grad av specialisering i AI.*

[21] ec.europe.eu/digital-single-market/en/news/digital-innovation-hubs-helping-companies-across-economy-make-most-digital-opportunities
[22] www.Ai4eu.eu.
[23] Europe.eu/investeu.

Digitala innovationsknutpunkter kan få stöd inom ramen för programmet för ett digitalt Europa.

- *Kommissionen och Europeiska investeringsfonden kommer att lansera ett pilotprojekt på 100 miljoner euro under första kvartalet 2020 för att tillhandahålla finansiering med eget kapital för innovativ AI-utveckling. Med förbehåll för en slutlig överenskommelse om den fleråriga budgetramen har kommissionen för avsikt att utöka det avsevärt från och med 2021 genom InvestEU.*

E. PARTNERSKAP MED DEN PRIVATA SEKTORN

Det är också mycket viktigt att se till att den privata sektorn deltar fullt ut i fastställandet av forsknings- och innovationsagendan och tillhandahåller den nivå av saminvestering som krävs. För detta krävs det att det upprättas ett brett offentlig-privat partnerskap och säkras ett åtagande från företagens högsta ledning.

- *Åtgärd 5: Inom ramen för Horisont Europa kommer kommissionen att inrätta ett nytt offentlig-privat partnerskap inom AI, data och robotteknik för att samarbeta, säkerställa samordning av forskning och innovation inom AI, samarbeta med andra offentlig-privata partnerskap i Horisont Europa och arbeta tillsammans med de testningsanläggningar och de digitala innovationsknutpunkter som nämns ovan.*

F. FRÄMJANDE AV DEN OFFENTLIGA SEKTORNS ANAMMANDE AV AI

Det är mycket viktigt att offentliga förvaltningar, sjukhus, allmännyttiga tjänsteföretag och transportföretag, finansiella tillsynsmyndigheter och andra aktörer på områden av allmänt intresse snabbt börjar använda produkter och AI-baserade tjänster i sin verksamhet. Särskild uppmärksamhet kommer att ägnas hälso- och sjukvård och transporter där tekniken är mogen för storskalig användning.

- *Åtgärd 6: Kommissionen kommer att inleda öppna och transparenta sektorsdialoger som prioriterar hälso- och sjukvård, förvaltningar i landsbygdsområden och offentliga tjänsteleverantörer, för att lägga fram en handlingsplan för att underlätta utveckling, experiment och införande. Sektorsdialogerna kommer att användas för att förbereda ett specifikt program för anammande av AI som stödjer offentlig upphandling av AI-system och bidrar till att förändra själva förfarandena för offentlig upphandling.*

G. SÄKRAD TILLGÅNG TILL DATA OCH INFRASTRUKTURER FÖR DATABEHANDLING

De åtgärdsområden som anges i den här vitboken kompletterar den plan som läggs fram parallellt inom ramen för EU:s datastrategi. Det är av grundläggande betydelse att tillgången till och hanteringen av data förbättras. Utan data är det inte möjligt att utveckla AI och andra digitala tillämpningar. Den enorma mängden nya data som ännu inte tagits fram utgör en möjlighet för EU att positionera sig i täten för den data- och AI-baserade omställningen. Att främja ansvarsfulla metoder för datahantering och datas överensstämmelse med Fair-principerna kommer att bidra till att bygga upp förtroende och säkerställa återanvändande av data[24]. Lika viktigt är investeringar i viktig datorteknik och datorinfrastruktur.

Kommissionen har föreslagit anslag på över 4 miljarder euro inom ramen för programmet för ett digitalt Europa för att stödja högpresterande datorsystem och kvantdatorsystem, inklusive perifer databehandling (*edge computing*) och AI, data och molninfrastruktur. De här prioriteringarna vidareutvecklas i EU:s datastrategi.

H. INTERNATIONELLA ASPEKTER

EU har goda förutsättningar att utöva globalt ledarskap när det gäller att skapa allianser kring gemensamma värden och främja etisk användning av AI. EU:s arbete med AI har redan påverkat de internationella diskussionerna. Vid utarbetandet av sina etiska riktlinjer involverade expertgruppen på hög nivå ett antal icke-EU-organisationer och flera statliga observatörer. Parallellt med detta var EU nära involverad i utarbetandet av OECD:s etiska principer för AI[25]. G20 godkände därefter de här principerna i ministeruttalandet om handel och digital ekonomi i juni 2019.

Samtidigt erkänner EU att viktigt arbete om AI ingår i andra multilaterala forum, bland annat Europarådet, Förenta nationernas organisation för utbildning, vetenskap och kultur (Unesco), Organisationen för ekonomiskt samarbete och utveckling (OECD), Världshandelsorganisationen (WTO) och Internationella teleunionen (ITU). I FN deltar EU i uppföljningen av rapporten från högnivåpanelen för digitalt samarbete (*High-Level Panel on Digital Cooperation*), inbegripet dess rekommendation om AI.

[24] Sökbara, tillgängliga, kompatibla och återanvändbara (FAIR, *Findable, Accessible, Interoperable and Reusable*), såsom anges i slutrapporten och handlingsplanen från kommissionens expertgrupp för FAIR-data, 2018, https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf.

[25] https://www.oecd.org/going-digital/ai/principles/

EU kommer att fortsätta att samarbeta med likasinnade länder, men även med globala aktörer, om AI, på grundval av en strategi som bygger på EU:s regler och värden (till exempel stöd till uppåtgående konvergens i lagstiftningen (*upward regulatory convergence*), tillgång till nyckelresurser, inklusive data, och skapande av lika villkor). Kommissionen kommer att noga övervaka politiken i tredjeländer som begränsar dataflöden och kommer att ta itu med otillbörliga begränsningar i bilaterala handelsförhandlingar och genom åtgärder inom ramen för Världshandelsorganisationen. Kommissionen är övertygad om att det internationella samarbetet kring AI måste bygga på en strategi som främjar respekten för de grundläggande rättigheterna, inbegripet mänsklig värdighet, mångfald, inkludering, icke-diskriminering och skydd av integritet och personuppgifter[26], och kommer att sträva efter att sprida sina värden i hela världen[27]. Det står också klart att ansvarsfull utveckling och användning av AI kan vara en drivkraft för att uppnå målen för hållbar utveckling och främja Agenda 2030.

5. ETT EKOSYSTEM AV FÖRTROENDE: REGELVERK FÖR AI

Som all ny teknik är användningen av AI förknippad med både möjligheter och risker. Medborgarna är rädda för att stå maktlösa i försvaret av sina rättigheter och sin säkerhet när de konfronteras med informationsasymmetrier i algoritmbaserade beslutsprocesser, och företagen är oroliga över den rättsliga osäkerheten. AI kan bidra till att skydda medborgarnas säkerhet och göra det möjligt för dem att åtnjuta sina grundläggande rättigheter, men medborgarna oroar sig också för att AI kan få oönskade effekter eller till och med användas i ont uppsåt. Den här oron måste tas på allvar. Brist på förtroende är dessutom, vid sidan av brist på investeringar och kompetens, en viktig faktor som hämmar en bredare spridning av AI.

Därför lade kommissionen den 25 april 2018 fram en AI-strategi[28] som behandlar de socioekonomiska aspekterna parallellt med ökade investeringar i forskning, innovation och AI-kapacitet i hela EU. Kommissionen enades med medlemsstaterna om en samordnad plan[29] för att anpassa strategierna. Kommissionen inrättade också en expertgrupp på hög nivå som offentliggjorde riktlinjer för tillförlitlig AI i april 2019[30].

Kommissionen offentliggjorde ett meddelande[31] där man välkomnade de sju centrala krav som fastställts i riktlinjerna från högnivågruppen:

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

[26] Inom ramen för partnerskapsinstrumentet kommer kommissionen att finansiera ett projekt på 2,5 miljoner euro som ska underlätta samarbetet med likasinnade partner för att främja EU:s etiska riktlinjer för AI och anta gemensamma principer och operativa slutsatser.
[27] Kommissionens ordförande Ursula von der Leyen, ”En ambitiösare union – Min agenda för Europa”, s. 17.
[28] COM(2018) 237.
[29] COM(2018) 795.
[30] https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top
[31] COM(2019) 168.

Riktlinjerna innehåller dessutom en bedömningslista som kan användas i praktiken av företag. Under andra halvåret 2019 har över 350 organisationer testat den här bedömningslistan och skickat in synpunkter. Högnivågruppen håller på att se över sina riktlinjer mot bakgrund av de här synpunkterna och kommer att slutföra det här arbetet senast i juni 2020. Ett viktigt resultat av den här återkopplingen är att även om ett antal av kraven redan återspeglas i befintliga rättsordningar eller regelverk, omfattas de krav som gäller transparens, spårbarhet och mänsklig uppsikt inte specifikt av den nuvarande lagstiftningen inom många ekonomiska sektorer.

Utöver den här uppsättningen icke-bindande riktlinjer från högnivågruppen, och i linje med kommissionsordförandens politiska riktlinjer, skulle ett tydligt EU-regelverk skapa förtroende för AI hos konsumenterna och företagen och därmed påskynda användandet av tekniken. Ett sådant regelverk bör vara förenligt med andra åtgärder för att främja Europas innovationskapacitet och konkurrenskraft på det här området. Dessutom måste det säkerställa socialt, miljömässigt och ekonomiskt optimala resultat och efterlevnad av EU:s lagstiftning, principer och värden. Detta är särskilt viktigt på områden där den direkta påverkan på medborgarnas rättigheter kan vara störst, till exempel när det gäller AI-tillämpningar för brottsbekämpning och rättsväsendet.

Utvecklare av AI och spridare av AI omfattas redan av EU-lagstiftningen om grundläggande rättigheter (till exempel skydd av personuppgifter, integritet, icke-diskriminering), konsumentskydd, produktsäkerhet och produktansvar. Konsumenterna förväntar sig samma säkerhetsnivå och respekt för sina rättigheter oavsett om en produkt eller ett system bygger på AI eller inte. Vissa särskilda aspekter av AI (till exempel bristande transparens) kan dock försvåra tillämpningen och kontrollen av efterlevnaden av den här lagstiftningen. Därför finns det ett behov av att undersöka om den nuvarande lagstiftningen kan hantera riskerna med AI och om kontrollen av dess efterlevnad kan ske effektivt, om det behövs anpassningar av lagstiftningen eller om det behövs ny lagstiftning.

Med tanke på hur snabbt AI utvecklas måste regelverket lämna utrymme för att ta hänsyn till den fortsatta utvecklingen. Eventuella ändringar bör begränsas till klart identifierade problem där det finns genomförbara lösningar.

Medlemsstaterna pekar på att det för närvarande saknas en gemensam europeisk ram. Den tyska kommittén för dataetik har efterlyst ett riskbaserat regleringssystem med fem nivåer som skulle gå från ingen reglering för de mest ofarliga AI-systemen till ett fullständigt förbud för de farligaste. Danmark har precis lanserat en prototyp på ett dataetiskt märke. Malta har infört ett frivilligt certifieringssystem för AI. Om EU misslyckas med att tillhandahålla en EU-omfattande strategi, finns det en reell risk för fragmentering av den inre marknaden, vilket skulle undergräva målen om förtroende, rättssäkerhet och marknadspenetration

Ett solitt europeiskt regelverk för tillförlitlig AI kommer att skydda alla EU-medborgare och bidra till att skapa en friktionsfri inre marknad för vidareutveckling och spridning av AI samt stärka EU:s industriella bas inom AI.

A. PROBLEMFORMULERING

Även om AI kan föra mycket gott med sig, till exempel genom att göra produkter och processer säkrare, kan den också göra skada. Den här skadan kan vara både materiell (människors säkerhet och hälsa, inbegripet förlust av människoliv, och sakskador) och immateriell (förlust av integritet, begränsningar av yttrandefriheten, mänsklig värdighet, diskriminering när det gäller till exempel tillträde till arbetsmarknaden) och kan gälla en mängd olika risker. Ett regelverk bör inriktas på hur man minimerar de olika potentiella skaderiskerna, särskilt de största.

De största riskerna i samband med användningen av AI rör tillämpningen av regler avsedda att skydda de grundläggande rättigheterna (inklusive skydd av personuppgifter och integritet samt icke-diskriminering) och frågor som rör säkerhet[32] och ansvar.

Risker för de grundläggande rättigheterna, inklusive skyddet för personuppgifter och integritet och icke-diskriminering

Användningen av AI kan påverka de värden som EU bygger på och leda till kränkningar av de grundläggande rättigheterna[33], inbegripet yttrandefriheten, mötesfriheten, rätten till mänsklig värdighet, rätten att inte utsättas för diskriminering på grund av kön, ras eller etniskt ursprung, religion eller övertygelse, funktionsnedsättning, ålder eller sexuell läggning, beroende på vad som är tillämpligt inom olika områden, rätten till skydd av personuppgifter och privatlivet, [34] eller rätten till ett effektivt rättsmedel och en opartisk domstol samt konsumentskydd. De här riskerna kan bero på brister i den allmänna utformningen av AI-systemen (inklusive mänsklig uppsikt) eller på användning av data utan korrigering av eventuella snedvridningar (till exempel systemet är tränat med hjälp av endast eller främst data som gäller män, vilket leder till suboptimala resultat när det är fråga om kvinnor).

AI kan utföra många av de funktioner som tidigare endast kunde utföras av människor. Till följd av detta kommer medborgare och juridiska personer i allt högre grad att bli föremål för åtgärder och beslut som fattas av eller med hjälp av AI-system, vilka ibland kan vara svåra att förstå och vid behov bestrida på ett effektivt sätt. Dessutom ökar AI möjligheterna att spåra och analysera människors vardagsvanor. Det kan till exempel finnas en risk för att AI, i strid med EU:s dataskyddsregler och andra regler, används av statliga myndigheter eller andra enheter för massövervakning och av arbetsgivare för att se hur deras anställda beter sig. Genom att analysera stora dataset och identifiera kopplingar mellan dem kan AI också användas för att spåra och avanonymisera data om personer, vilket skapar nya risker för skydd av personuppgifter även i fråga om dataset som i sig inte innehåller personuppgifter. AI används också av mellanhänder på internet för att prioritera information åt sina användare och redigera innehållet. De behandlade uppgifterna, tillämpningarnas utformning och utrymmet för mänskligt ingripande kan påverka yttrandefriheten, skyddet av personuppgifter, integriteten och de politiska friheterna.

[32] Detta inbegriper frågor om cybersäkerhet, frågor i samband med AI-tillämpningar i kritisk infrastruktur, eller användning av AI i ont uppsåt.

[33] Europarådets forskning visar att användningen av AI kan få konsekvenser för ett stort antal grundläggande rättigheter, https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5.

[34] Den allmänna dataskyddsförordningen och direktivet om integritet och elektronisk kommunikation (en ny förordning om integritet och elektronisk kommunikation är under utarbetande) tar upp de här riskerna, men det kan finnas behov av att undersöka om AI-system medför ytterligare risker. Kommissionen kommer kontinuerligt att övervaka och utvärdera tillämpningen av den allmänna dataskyddsförordningen.

Vissa AI-algoritmer kan, när de används för att förutsäga återfall i brott, uppvisa snedvridning i fråga om kön och ras, och visa olika sannolikhet för återfall i brott för kvinnor jämfört med män eller för medborgare i ett land jämfört med utländska medborgare. *Källa: Tolan S., Miron M., Gomez E. and Castillo C. Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia, Best Paper Award, International Conference on AI and Law, 2019.*

Vissa AI-program för ansiktsanalys visar snedvridning i fråga om kön och ras, och uppvisar liten felfrekvens vid bestämning av ljushyade mäns kön men stor felfrekvens vid könsbestämning av mörkhyade kvinnor. Källa: *Joy Buollamwini, Timnit Gebru; Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:77-91, 2018*.

Snedvridning och diskriminering är en inneboende risk vid all samhällelig eller ekonomisk verksamhet. Mänskligt beslutsfattande är inte immunt mot misstag och snedvridning. Samma snedvridning vid användande av AI skulle dock kunna få mycket större effekt och påverka och diskriminera många människor i frånvaro av de mekanismer för social kontroll som styr människors beteende[35]. Detta kan också inträffa när AI-system ”lär sig” medan de används. I sådana fall, där resultatet inte kan förhindras eller förutses i konstruktionsfasen, kommer riskerna inte att härröra från ett fel i systemets ursprungliga konstruktion, utan från de praktiska effekterna av de korrelationer eller mönster som systemet identifierar i ett stort dataset.

De särskilda egenskaper som ofta finns hos AI-tekniken, bland annat bristande transparens (”svarta lådan-effekt”), komplexitet, oförutsägbarhet och delvis autonomt beteende, kan göra det svårt att kontrollera överensstämmelsen med, och hindra effektiv kontroll av efterlevnaden av, regler i befintlig EU-lagstiftning avsedda att skydda de grundläggande rättigheterna. Tillsynsmyndigheterna och berörda personer kan sakna möjlighet att kontrollera hur ett visst beslut fattades med hjälp av AI och därmed huruvida de relevanta reglerna följdes. Enskilda och juridiska personer kan mötas av svårigheter att få tillgång till faktisk rättslig prövning i situationer där sådana beslut kan ha påverkat dem negativt.

Risker för säkerheten och ett effektivt fungerande ansvarssystem

AI-teknik kan innebära nya säkerhetsrisker för användare när den är inbäddad i produkter och tjänster. Till exempel kan en självkörande bil, till följd av en brist i tekniken för att känna igen föremål, felaktigt identifiera ett föremål på vägen och orsaka en olycka med personskador och materiella skador. När det gäller riskerna för de grundläggande rättigheterna kan de orsakas av brister i utformningen av AI-tekniken eller bero på problem med tillgången till och kvaliteten på data, eller andra problem som orsakas av maskininlärning. Vissa av de här riskerna är inte begränsade till produkter och tjänster som bygger på AI, men användningen av AI kan öka eller förvärra riskerna.

[35] Kommissionens rådgivande kommitté för lika möjligheter för kvinnor och män håller för närvarande på att utarbeta ett ”Yttrande om artificiell intelligens” som rymmer en analys av bland annat effekterna av artificiell intelligens för jämställdhet mellan kvinnor och män, och som förväntas antas av kommittén i början av 2020. EU-strategin för jämställdhet 2020–2024 tar också upp kopplingen mellan AI och jämställdhet mellan kvinnor och män. Det europeiska nätverket för likabehandlingsorgan (Equinet) kommer att publicera rapporten (författad av Robin Allen och Dee Masters) *Regulating AI: the new role for Equality Bodies – Meeting the new challenges to equality and non-discrimination from increased digitalisation and the use of AI*, som väntas i början av 2020.

Brist på tydliga säkerhetsbestämmelser för att hantera de här riskerna kan, utöver riskerna för de berörda personerna, skapa rättslig osäkerhet för företag som inom EU marknadsför produkter som innefattar AI. De myndigheter som ansvarar för marknadsövervakning och kontroll av reglernas efterlevnad kan hamna i en situation där de är osäkra på om de kan ingripa, eftersom de kanske inte har befogenhet att agera och/eller inte har lämplig teknisk kapacitet att inspektera systemen[36]. Rättslig osäkerhet kan därför minska säkerheten totalt sett och undergräva de europeiska företagens konkurrenskraft.

Om säkerhetsriskerna skulle uppstå, gör bristen på tydliga krav och egenskaperna hos den ovannämnda AI-tekniken det svårt att spåra potentiellt problematiska beslut som fattas med stöd av AI-system. Detta kan i sin tur göra det svårt för personer som lidit skada att få ersättning enligt befintlig ansvarslagstiftning på nationell nivå och EU-nivå[37].

Enligt direktivet om produktansvar är en tillverkare ansvarig för skador som orsakas av en felaktig produkt. När det gäller ett AI-baserat system såsom självkörande bilar kan det dock vara svårt att bevisa att produkten är felaktig, att skadan uppstått och att det finns ett orsakssamband mellan dem. Det råder dessutom viss osäkerhet om hur och i vilken utsträckning direktivet om produktansvar är tillämpligt på vissa typer av fel, till exempel om de har uppkommit på grund av svagheter i produktens cybersäkerhet.

Svårigheten att spåra potentiellt problematiska beslut som fattats av AI-system och avses ovan beträffande grundläggande rättigheter gäller således även frågor som rör säkerhet och ansvar. Personer som lidit skada har eventuellt inte faktisk tillgång till den bevisning som krävs för att driva ett ärende i domstol, till exempel, och har eventuellt mindre faktiska möjligheter till rättslig prövning jämfört med situationer där skadan orsakas av traditionell teknik. De här riskerna kommer att öka i takt med att användningen av AI blir vanligare.

B. EVENTUELLA ANPASSNINGAR TILL EU:S BEFINTLIGA RÄTTSLIGA RAM FÖR AI

Ett omfattande befintligt EU-regelverk för produktsäkerhet och produktansvar[38], inklusive sektorsspecifika regler, som kompletteras av nationell lagstiftning, är relevant för och eventuellt tillämpligt på ett antal framväxande AI-tillämpningar.

[36] Ett exempel på detta kan vara smartklockan för barn. Den här produkten vållar inte det barn som bär den någon direkt skada, men eftersom den inte erbjuder säkerhet ens på miniminivå kan den lätt användas för att få tillgång till barnet. De myndigheter som ansvarar för marknadsövervakning kan få svårt att ingripa i de fall då risken inte är kopplad till produkten som sådan.

[37] Konsekvenserna av AI, sakernas internet och annan digital teknik för säkerhets- och ansvarslagstiftning analyseras i den kommissionsrapport som åtföljer den här vitboken.

[38] EU:s rättsliga ram för produktsäkerhet består av direktivet om allmän produktsäkerhet (direktiv 2001/95/EG), som skyddsnät, och ett antal sektorsspecifika regler som omfattar olika produktkategorier, från maskiner, flygplan och bilar till

När det gäller skyddet för de grundläggande rättigheterna och konsumenträttigheterna omfattar EU-regelverket lagstiftning såsom direktivet om likabehandling oavsett ras[39], direktivet om likabehandling i arbetslivet[40], direktiven om likabehandling av kvinnor och män när det gäller anställning och tillgång till varor och tjänster[41], ett antal konsumentskyddsregler[42], liksom regler om skydd av personuppgifter och integritet, i synnerhet den allmänna dataskyddsförordningen och annan sektorslagstiftning som omfattar skydd av personuppgifter, såsom dataskyddsdirektivet för brottsbekämpning[43]. Vidare kommer, från och med 2025, de bestämmelser om tillgänglighetskrav för varor och tjänster som anges i den europeiska rättsakten om tillgänglighet att vara gällande[44]. Dessutom måste de grundläggande rättigheterna respekteras vid genomförandet av annan EU-lagstiftning, bland annat när det gäller finansiella tjänster, migration eller ansvar för mellanhänder på internet.

Även om EU-lagstiftningen fortfarande i princip är fullt tillämplig oavsett involveringen av AI, är det viktigt att bedöma om den kan genomföras på ett adekvat sätt för att hantera de risker som AI-system skapar, eller om det behövs anpassningar av specifika rättsliga instrument.

Exempelvis förblir ekonomiska aktörer fullt ansvariga för AI:s överensstämmelse med befintliga konsumentskyddsregler, och det ska inte vara tillåtet att använda något algoritmiskt utnyttjande av konsumenternas beteende i strid med befintliga regler, och brott mot de här reglerna ska bestraffas.

Kommissionen anser att den rättsliga ramen skulle kunna förbättras för att hantera följande risker och situationer:

- *Effektiv tillämpning och kontroll av efterlevnad av befintlig EU-lagstiftning och nationell lagstiftning*: de viktigaste egenskaperna hos AI skapar problem när det gäller att säkerställa korrekt tillämpning och efterlevnad av EU-lagstiftning och nationell lagstiftning. Bristen transparens hos AI gör det svårt att identifiera och bevisa eventuella överträdelser av lagar, inklusive lagbestämmelser som skyddar de grundläggande rättigheterna, att fastställa ansvar och att uppfylla villkoren för att begära ersättning. För att säkerställa en effektiv tillämpning och efterlevnad kan det därför vara nödvändigt att anpassa eller förtydliga den befintliga lagstiftningen på vissa områden, till exempel i fråga om ansvar enligt den närmare beskrivningen i den rapport som åtföljer den här vitboken.

- *Begränsningar av tillämpningsområdet för gällande EU-lagstiftning*: en viktig del av EU:s lagstiftning om produktsäkerhet gäller utsläppande av produkter på marknaden. Även om programvara, när den ingår i slutprodukten, måste överensstämma med de relevanta produktsäkerhetsreglerna i EU:s lagstiftning om produktsäkerhet, är det en öppen fråga om fristående programvara omfattas av EU:s produktsäkerhetslagstiftning, utanför vissa sektorer med uttryckliga regler[45]. De allmänna EU-säkerhetsbestämmelserna som är i kraft i dag gäller

leksaker och medicintekniska produkter, i syfte att ge en hög nivå av hälsa och säkerhet. Lagstiftningen om produktansvar kompletteras av olika system för civilrättsligt ansvar för skador som orsakas av produkter eller tjänster.

[39] Direktiv 2000/43/EG.

[40] Direktiv 2000/78/EG.

[41] Direktiv 2004/113/EG, direktiv 2006/54/EG.

[42] Till exempel direktivet om otillbörliga affärsmetoder (direktiv 2005/29/EG) och direktivet om konsumenträttigheter (direktiv 2011/83/EG).

[43] Europaparlamentets och rådets direktiv (EU) 2016/680 av den 27 april 2016 om skydd för fysiska personer med avseende på behöriga myndigheters behandling av personuppgifter för att förebygga, förhindra, utreda, avslöja eller lagföra brott eller verkställa straffrättsliga påföljder, och det fria flödet av sådana uppgifter.

[44] Direktiv (EU) 2019/882 om tillgänglighetskrav för produkter och tjänster.

[45] Till exempel anses programvara som av tillverkaren avses att användas för medicinska ändamål vara en medicinteknisk produkt enligt förordningen om medicintekniska produkter (förordning (EU) 2017/745).

för produkter, inte tjänster, och därför i princip inte heller tjänster som bygger på AI-teknik (till exempel hälso- och sjukvårdstjänster, finansiella tjänster och transporttjänster).

- *Förändrad funktionalitet hos AI-system*: integreringen av programvara, inklusive AI, i produkter kan ändra de här produkternas och systemens funktion under deras livscykel. Detta gäller särskilt för system som kräver regelbundna uppdateringar av programvara eller som bygger på maskininlärning. De här egenskaperna kan ge upphov till nya risker som inte fanns när systemet släpptes ut på marknaden. De här riskerna behandlas inte i tillräcklig utsträckning i den befintliga lagstiftningen som främst är inriktad på de säkerhetsrisker som förelåg vid tidpunkten för utsläppandet på marknaden.

- *Osäkerhet rörande ansvarsfördelningen mellan olika ekonomiska aktörer i leveranskedjan*: i EU:s lagstiftning om produktsäkerhet åligger i allmänhet ansvaret tillverkaren av den produkt som släpps ut på marknaden, inbegripet alla komponenter, till exempel AI-system. Men reglerna kan till exempel bli oklara, om AI läggs till av en annan part än tillverkaren efter det att produkten släppts ut på marknaden. EU:s lagstiftning om produktansvar föreskriver dessutom tillverkarnas ansvar och överlåter åt de nationella ansvarsreglerna att reglera ansvaret för andra aktörer i leveranskedjan.

- *Ändringar av begreppet säkerhet*: användningen av AI i produkter och tjänster kan ge upphov till risker som för närvarande inte uttryckligen behandlas i EU-lagstiftningen. De här riskerna kan vara kopplade till cyberhot, personliga säkerhetsrisker (som till exempel kan kopplas till nya tillämpningar av AI, till exempel i hushållsapparater), risker till följd av förlust av uppkopplingsmöjlighet och så vidare. De här riskerna kan föreligga vid tidpunkten för produktens utsläppande på marknaden eller uppstå till följd av programvaruuppdateringar eller självinlärande när produkten används. EU bör till fullo utnyttja de verktyg som står till dess förfogande för att stärka sitt faktaunderlag om potentiella risker i samband med AI-tillämpningar, bland annat med hjälp av erfarenheterna från EU:s cybersäkerhetsbyrå (Enisa) för bedömning av AI-hotbilden.

Som tidigare nämnts undersöker flera medlemsstater redan olika alternativ för nationell lagstiftning att hantera de utmaningar som AI medför. Detta ökar risken för fragmentering av den inre marknaden. Nationella regler som avviker från varandra kommer sannolikt att skapa hinder för företag som vill sälja och driva AI-system på den inre marknaden. En gemensam strategi på EU-nivå skulle göra det möjligt för europeiska företag att få smidigt tillträde till den inre marknaden och skulle stödja deras konkurrenskraft på de globala marknaderna.

Rapport om konsekvenser för säkerhet och ansvar när det gäller AI, sakernas internet och robotteknik

Rapporten, som åtföljer den här vitboken, innehåller en analys av den relevanta rättsliga ramen. Den identifierar oklarheter när det gäller tillämpningen av den här ramen med avseende på de specifika risker som förorsakas av AI-system och annan digital teknik.

Slutsatsen dras att den nuvarande produktsäkerhetslagstiftningen redan stöder ett utvidgat säkerhetsbegrepp som skyddar mot alla typer av risker i förbindelse med produktens användning. Bestämmelser som uttryckligen täcker nya risker som uppstår genom den framväxande digitala tekniken skulle emellertid kunna införas för att skapa ökad rättslig säkerhet.

- Det autonoma beteendet hos vissa AI-system under deras livscykel kan medföra viktiga produktförändringar som påverkar säkerheten, vilket kan kräva en ny riskbedömning. Mänsklig tillsyn, från AI-produkternas och AI-systemens konstruktionsfas och under hela livscykeln kan dessutom behövas som en skyddsåtgärd.
- Uttryckliga skyldigheter för tillverkarna skulle kunna övervägas även i fråga om psykiska säkerhetsrisker för användare där så är lämpligt (till exempel vid samarbete med människoliknande robotar).
- Unionens lagstiftning om produktsäkerhet skulle kunna föreskriva särskilda krav för att hantera säkerhetsrisker förknippade med felaktiga data på konstruktionsstadiet samt mekanismer för att säkerställa att kvaliteten på data upprätthålls under hela användningen av AI-produkter och AI-system.
- Den bristande transparensen hos system baserade på algoritmer kan hanteras genom krav på transparens.
- Befintliga regler kan behöva anpassas och förtydligas när det gäller en fristående programvara som påverkar säkerheten, när den själv släpps ut på marknaden eller när den laddas ner till en produkt efter det att den släppts ut på marknaden.
- Med tanke på den ökande komplexiteten i leveranskedjorna vad gäller ny teknik skulle bestämmelser som specifikt kräver samarbete mellan de ekonomiska aktörerna i leveranskedjan och användarna kunna ge rättssäkerhet.

Egenskaperna hos framväxande digital teknik, såsom AI, sakernas internet och robotteknik, kan utmana vissa aspekter av ansvarsramarna och minska deras effektivitet. Vissa av de här egenskaperna skulle kunna göra det svårt att spåra skadan tillbaka till en person, vilket skulle vara nödvändigt för ett culpabaserat krav i enlighet med de flesta nationella bestämmelser. Detta kan avsevärt öka kostnaderna de målsägande som innebär att skadeståndsanspråk mot andra än tillverkaren kan vara svåra att göra eller bevisa.

- Personer som har lidit skada som orsakats av AI-system måste åtnjuta samma skydd som personer som har lidit skada som orsakats av annan teknik, samtidigt som fortsatt utveckling av teknisk innovation bör tillåtas.
- Alla alternativ för att säkerställa det här målet bör noggrant bedömas, inbegripet eventuella ändringar av direktivet om produktansvar och en eventuell ytterligare harmonisering av de nationella ansvarsbestämmelserna. Kommissionen vill till exempel inhämta synpunkter på om, och i vilken utsträckning, detta kan behövas för att mildra konsekvenserna av komplexiteten genom att anpassa

Av diskussionen ovan drar kommissionen slutsatsen att – utöver eventuella anpassningar av befintlig lagstiftning – en ny särskild lagstiftning om AI kan behövas för att anpassa EU:s rättsliga ram till den aktuella och förväntade tekniska och kommersiella utvecklingen.

C. TILLÄMPNINGSOMRÅDET FÖR EU:S FRAMTIDA REGELVERK

En nyckelfråga för det framtida särskilda regelverket för AI är att fastställa dess tillämpningsområde. Arbetshypotesen är att regelverket ska gälla produkter och tjänster som bygger på AI. AI bör därför tydligt definieras i den här vitboken, liksom i alla eventuella framtida politiska initiativ.

Kommission gav i sitt meddelande ”Artificiell intelligens för Europa” en första definition av AI[46]. Den definitionen preciserades ytterligare av expertgruppen på hög nivå[47].

I ett nytt rättsligt instrument måste definitionen av AI vara tillräckligt flexibel för att kunna anpassas till den tekniska utvecklingen och samtidigt vara tillräckligt exakt för att ge den rättssäkerhet som krävs.

Det förefaller viktigt att i den här vitboken, och i alla eventuella framtida diskussioner om politiska initiativ, klargöra de huvudsakliga delar som AI utgörs av, nämligen ”data” och ”algoritmer”. AI kan integreras i maskinvara. När det gäller maskininlärningsmetoder, som ingår i AI, tränas algoritmer att utifrån en uppsättning data härleda vissa mönster för att avgöra vilka åtgärder som krävs för att uppnå ett visst mål. Algoritmer kan fortsätta att lära sig när de används. Även om AI-baserade produkter kan agera självständigt genom att uppfatta sin omgivning och utan att följa en förutbestämd uppsättning instruktioner, är beteendet till stor del definierat och begränsat av utvecklaren. Människor fastställer och programmerar de mål som ett AI-system ska optimeras för.

> I självkörande fordon, till exempel, använder sig algoritmen i realtid av data från bilen (hastighet, motorns bränsleförbrukning, stötdämpare m.m.) och från sensorer som känner av bilens hela omgivning (vägen, vägskyltar, andra fordon, fotgängare osv.) så att den kan härleda vilken riktning, acceleration och hastighet bilen bör ha för att nå en viss destination. Baserat på de data som observeras anpassar sig algoritmen till vägförhållandena och till andra yttre förhållanden, till exempel andra förares beteende, för att härleda det säkraste och bekvämaste körsättet.

EU har ett strikt regelverk för att bland annat garantera konsumentskyddet, ta itu med otillbörliga affärsmetoder och säkerställa skyddet av personuppgifter och integritet. Dessutom innehåller det gemensamma EU-regelverket särskilda regler för vissa sektorer (till exempel hälso- och sjukvård och transporter). De här befintliga bestämmelserna i EU-lagstiftningen kommer att fortsätta att gälla för AI, även om vissa uppdateringar av regelverket kan vara nödvändiga för att återspegla den digitala omställningen och användningen av AI (se avsnitt B). Följaktligen kommer de aspekter som redan

[46] COM(2018) 237 final, s. 1: ”Artificiell intelligens avser system som uppvisar intelligent beteende genom att analysera sin miljö och vidta åtgärder – med viss grad av självständighet – för att uppnå särskilda mål.
AI-baserade system kan vara helt programvarubaserade och fungera i den virtuella världen (till exempel röstassistenter, bildanalysprogram, sökmotorer, tal- och ansiktsigenkänningssystem), eller inbäddas i maskinvaruenheter (till exempel avancerade robotar, självkörande bilar, drönare eller tillämpningar för sakernas internet).”

[47] Expertgruppen på hög nivå, en definition av AI, s. 8: ”Artificiella intelligenssystem (AI) är programvarusystem (och eventuellt även maskinvarusystem) som har konstruerats av människor och som när de får ett komplext mål agerar i den fysiska eller digitala dimensionen genom att uppfatta sin omgivning via datainsamling och att tolka insamlade strukturerade eller ostrukturerade data, resonerar om den kunskap eller behandlar den information som härletts ur den här datan och beslutar om den bästa åtgärd eller de bästa åtgärderna som ska vidtas för att uppnå det fastställda målet. AI-system kan använda symboliska regler eller lära sig en numerisk modell. De kan också anpassa sitt beteende genom att analysera hur miljön har påverkats av deras föregående åtgärder.”

omfattas av befintlig horisontell eller sektoriell lagstiftning (till exempel om medicintekniska produkter[48] och inom transportsystem) även fortsättningsvis att omfattas av den här lagstiftningen.

Av principskäl bör det nya regelverket för AI vara effektivt för att uppnå sina mål, samtidigt som det inte får bli alltför föreskrivande så att det skapar en oproportionerlig börda, särskilt för små och medelstora företag. För att uppnå den här balansen anser kommissionen att den bör följa en riskbaserad metod.

En riskbaserad metod är viktig för att säkerställa att regleringsåtgärderna är proportionerliga. Det krävs dock tydliga kriterier för att skilja mellan de olika AI-tillämpningarna, särskilt huruvida det är tillämpningar ”med hög risk” eller inte[49]. Vad som anses utgöra en AI-tillämpning med hög risk bör vara tydligt och lättförståeligt och gälla för alla berörda parter. Emellertid omfattas även AI-tillämpningar som inte anses medföra hög risk helt av de befintliga EU-reglerna.

Kommissionen anser att en given AI-tillämpning generellt bör betraktas som en tillämpning med hög risk mot bakgrund av vad som står på spel, med beaktande av huruvida både sektorn och den avsedda användningen innebär betydande risker, i synnerhet när det gäller att skydda säkerheten, konsumenternas rättigheter och de grundläggande rättigheterna. Närmare bestämt bör en AI-tillämpning betraktas som en tillämpning med hög risk om den uppfyller följande två kumulativa kriterier:

- För det första, AI-tillämpningen används i en sektor där betydande risker kan förväntas uppstå med tanke på de typiska egenskaperna hos den verksamhet som bedrivs. Detta första kriterium säkerställer att regleringsåtgärderna inriktas på de områden där det generellt bedöms mest sannolikt att risker uppstår. De sektorer som omfattas bör förtecknas särskilt och uttömmande i det nya regelverket, till exempel hälso- och sjukvård, transport, energi och delar av den offentliga sektorn[50]. Förteckningen bör ses över regelbundet och, när så är nödvändigt, ändras i enlighet med den relevanta utvecklingen av praxis.

- För det andra, AI-tillämpningen inom sektorn i fråga används dessutom på ett sådant sätt att betydande risker sannolikt uppstår. Detta andra kriterium återspeglar det faktum att inte varje användning av AI i de utvalda sektorerna medför betydande risker. Även om hälso- och sjukvården generellt kan vara en relevant sektor, utgör till exempel en brist i tidsbokningssystemet på ett sjukhus normalt inte en risk av sådan betydelse att det motiverar lagstiftningsåtgärder. Bedömningen av risknivån för en viss användning skulle kunna baseras på effekterna för de berörda parterna, till exempel användning av AI-tillämpningar som får rättsliga eller andra betydande effekter för en privatpersons eller ett företags rättigheter, som utgör en risk för personskador, dödsfall eller betydande materiella eller immateriella skador, eller som får effekter som inte rimligen kan undvikas av fysiska eller juridiska personer.

Om de här två kumulativa kriterierna tillämpas kommer det att säkerställa att regelverkets tillämpningsområde är riktat och skapar rättslig säkerhet. De obligatoriska kraven i det nya regelverket

[48] Till exempel finns det olika säkerhetsaspekter och rättsliga konsekvenser för AI-system som tillhandahåller specialiserad medicinsk information till läkare, AI-system som tillhandahåller medicinsk information direkt till patienter och AI-system som själva utför medicinska uppgifter direkt på en patient. Kommissionen undersöker de här säkerhets- och ansvarsproblemen som är specifika för hälso- och sjukvård.

[49] EU-lagstiftningen kan kategorisera ”risker” annorlunda än vad som beskrivs här, beroende på området, till exempel i fråga om produktsäkerhet.

[50] I den offentliga sektorn skulle områden såsom asyl, migration, gränskontroller, rättsväsendet, social trygghet och arbetsförmedling kunna omfattas.

för AI (se avsnitt D nedan) skulle i princip endast gälla de tillämpningar som identifieras som tillämpningar med hög risk i enlighet med de här två kumulativa kriterierna.

Trots vad som sägs ovan kan det finnas undantagsfall där, mot bakgrund av de aktuella riskerna, användningen av AI-tillämpningar för vissa ändamål ska betraktas som hög risk i sig, det vill säga oavsett sektor som berörs, och där nedanstående krav fortfarande skulle gälla[51]. För att åskådliggöra detta skulle man särskilt kunna föreställa sig följande:

- Användningen av AI-tillämpningar för rekryteringsprocesser och i situationer som påverkar arbetstagares rättigheter skulle, med tanke på betydelsen för enskilda personer och EU:s regelverk om likabehandling i arbetslivet, alltid betraktas som användning med ”hög risk”, och därför skulle kraven nedan alltid gälla. Ytterligare särskilda tillämpningar som påverkar konsumenters rättigheter skulle kunna övervägas.

- Användning av AI-tillämpningar för biometrisk fjärridentifiering[52] och annan inkräktande övervakningsteknik skulle alltid anses utgöra användning med ”hög risk” och därför skulle kraven nedan alltid gälla.

D. TYPER AV KRAV

Vid utformningen av det framtida regelverket för AI kommer det att bli nödvändigt att besluta om vilka typer av obligatoriska rättsliga krav som ska ställas på de berörda aktörerna. De kraven kan specificeras ytterligare genom standarder. Såsom konstateras i avsnitt C ovan och utöver nu befintliga lagstiftning, skulle de kraven endast gälla AI-tillämpningar med hög risk, och på så sätt säkerställa att eventuella regleringsåtgärder är fokuserade och proportionerliga.

Med beaktande av högnivågruppens riktlinjer och vad som anförts ovan kan kraven för AI-tillämpningar med hög risk bestå av följande centrala aspekter, vilka diskuteras mer ingående i underavsnitten nedan:

- Träningsdata.
- Data och registerföring.
- Tillhandahållande av information.
- Robusthet och korrekthet.
- Mänsklig tillsyn.
- Särskilda krav för vissa specifika AI-tillämpningar, till exempel de som används för biometrisk fjärridentifiering.

För att garantera rättslig säkerhet kommer de här kraven att specificeras ytterligare för att ge ett tydligt riktmärke för alla aktörer som måste följa dem.

a) Träningsdata

[51] Det är viktigt att betona att även annan EU-lagstiftning kan vara tillämplig. Exempelvis kan direktivet om allmän produktsäkerhet omfatta säkerheten hos AI-tillämpningar när de ingår i en konsumentprodukt.

[52] Biometrisk fjärridentifiering bör skiljas från biometrisk autentisering (det senare är en säkerhetsprocess som bygger på en individs unika biologiska egenskaper för att verifiera att individen är den som den uppger sig vara). Med biometrisk fjärridentifiering avses när flera personers identitet fastställs med hjälp av biometriska kännetecken (fingeravtryck, ansiktsbild, iris, venmönster osv.) på distans, på en offentlig plats och på ett kontinuerligt eller pågående sätt genom att jämföra dem med uppgifter som lagras i en databas.

Det är viktigare än någonsin att främja, stärka och försvara EU:s värden och regler, särskilt de rättigheter som medborgarna har enligt EU-rätten. De här insatserna omfattar utan tvivel även de AI-tillämpningar med hög risk som marknadsförs och används i EU, och som det här är fråga om.

Som tidigare diskuterats finns det ingen AI utan data. Det sätt på vilket många AI-system fungerar, och vilka åtgärder och beslut de kan leda till, beror i hög grad på de data som systemen har tränats på. Därför bör nödvändiga åtgärder vidtas för att se till att EU:s värden och regler respekteras när det gäller de data som används för att träna AI-system, särskilt i fråga om säkerhet och gällande lagstiftning för skydd av de grundläggande rättigheterna. Följande krav avseende data som används för att träna AI-system kan förutses:

- Krav som syftar till att ge rimliga garantier för att den efterföljande användningen av de produkter eller tjänster som AI-systemet möjliggör är säker, genom att det uppfyller de normer som fastställs i EU:s befintliga säkerhetsbestämmelser (både befintliga och eventuella kompletterande), till exempel krav som säkerställer att AI-system tränas på dataset som är tillräckligt breda och täcker alla relevanta scenarier som behövs för att undvika farliga situationer.

- Krav på att vidta rimliga åtgärder för att säkerställa att sådan senare användning av AI-system inte leder till resultat som medför förbjuden diskriminering. De här kraven kan särskilt medföra skyldigheter att använda dataset som är tillräckligt representativa, särskilt för att säkerställa att alla relevanta aspekter av kön, etnicitet och andra möjliga grunder för förbjuden diskriminering återspeglas i de här dataseten på lämpligt sätt.

- Krav som syftar till att säkerställa att integritet och personuppgifter skyddas på ett adekvat sätt vid användning av AI-baserade produkter och tjänster. De frågor som omfattas av den allmänna dataskyddsförordningen och dataskyddsdirektivet för brottsbekämpning regleras av de här rättsakterna.

b) Registerföring och datalagring

Med beaktande av faktorer såsom komplexiteten hos och bristen på transparens hos många AI-system och därmed sammanhängande svårigheter att effektivt kontrollera överenstämmelsen med och efterlevnaden av de tillämpliga reglerna, måste krav ställas på registerföring i samband med programmeringen av algoritmen, de data som används för att träna AI-system med hög risk och, i vissa fall, lagring av själva datasetet. De här kraven gör det i princip möjligt att spåra och kontrollera sådana åtgärder och beslut av AI-system som är potentiellt problematiska. Detta skulle inte endast underlätta tillsyn och kontroll av efterlevnad, utan även öka incitamenten för de berörda ekonomiska aktörerna att i ett tidigt skede beakta behovet av att följa de här reglerna.

För att åstadkomma detta kan det i regelverket föreskrivas att följande ska dokumenteras:

- Korrekta register över rörande det dataset som används för att träna och testa AI-systemen, inklusive en beskrivning av de huvudsakliga egenskaperna och hur datasetet har valts ut.

- Dataseten, i vissa motiverade fall.

- Metoder, processer och tekniker för programmering[53] och träning som används för att bygga, testa och validera AI-systemen, inbegripet, i förekommande fall, med avseende på säkerhet och undvikande av snedvridning som skulle kunna leda till förbjuden diskriminering,

Registren, dokumentationen och, i förekommande fall, dataseten skulle behöva bevaras under en begränsad och rimligt lång tidsperiod för att säkerställa en effektiv kontroll av efterlevnaden av den relevanta lagstiftningen. Åtgärder bör vidtas för att se till att de görs tillgängliga på begäran, särskilt för behöriga myndigheters testning eller inspektion. Vid behov bör åtgärder vidtas för att se till att konfidentiell information, såsom företagshemligheter, skyddas.

c) Tillhandahållande av information

Transparens krävs också utöver de krav på registerföring som diskuteras i punkt c ovan. För att uppnå de eftersträvade målen – särskilt att främja en ansvarsfull användning av AI, skapa förtroende och underlätta rättslig prövning vid behov – är det viktigt att det på ett proaktivt sätt tillhandahålls adekvat information om användningen av AI-system med hög risk.

Därför skulle följande krav kunna övervägas:

- Säkerställa att tydlig information om AI-systemens kapacitet och begränsningar tillhandahålls, särskilt deras avsedda ändamål, de förhållanden under vilka de kan förväntas fungera såsom avsett och den förväntade nivån av korrekthet när det gäller att uppnå det angivna syftet. Den här informationen är viktig, särskilt för spridare av systemen, men den kan också vara relevant för behöriga myndigheter och berörda parter.

- Vid sidan av detta bör medborgarna få tydlig information om när de samverkar med ett AI-system och inte en människa. EU:s dataskyddslagstiftning innehåller redan vissa bestämmelser av det här slaget[54], men ytterligare krav kan behövas för att uppnå de ovannämnda målen. Om så är fallet bör onödiga bördor undvikas. Därför behöver sådan information till exempel inte lämnas i situationer där det är direkt uppenbart för medborgarna att de interagerar med AI-system. Det är vidare viktigt att den information som tillhandahålls är objektiv, koncis och lättförståelig. Det sätt på vilket informationen ska lämnas bör anpassas till det specifika sammanhanget.

d) Robusthet och korrekthet

AI-system – särskilt AI-tillämpningar med hög risk – måste vara tekniskt robusta och korrekta för att vara tillförlitliga. Det innebär att sådana system måste utvecklas på ett ansvarsfullt sätt och under vederbörligt övervägande i förväg av de risker som de kan medföra. De måste utvecklas och fungera så att det säkerställs att AI-systemen fungerar tillförlitligt så som de är avsedda att göra. Alla rimliga åtgärder bör vidtas för att minimera risken för att skada uppstår.

Följaktligen skulle följande faktorer kunna övervägas:

- Krav som säkerställer att AI-systemen är robusta och korrekta, eller åtminstone korrekt återspeglar sin grad av korrekthet under alla faser av livscykeln.

[53] till exempel om algoritmen, inklusive vad modellen ska optimeras för, vilka vikter som inledningsvis ska tilldelas vissa parametrar och så vidare.

[54] I synnerhet ska registeransvariga, i enlighet med artikel 13.2 f i den allmänna dataskyddsförordningen, vid tidpunkten då personuppgifterna erhålls, ge de registrerade ytterligare information om förekomsten av automatiserat beslutsfattande som är nödvändig för att säkerställa en rättvis och transparent behandling, samt viss ytterligare information.

- Krav som säkerställer att resultaten är reproducerbara.

- Krav som säkerställer att AI-system kan hantera fel eller inkonsekvenser på ett tillfredsställande sätt under alla faser av livscykeln.

- Krav som säkerställer att AI-system är motståndskraftiga mot såväl uppenbara angrepp som mer subtila försök att manipulera data eller själva algoritmerna, och att riskreducerande åtgärder vidtas i sådana fall.

e) Mänsklig tillsyn

Mänsklig tillsyn bidrar till att säkerställa att ett AI-system inte undergräver människors självständighet eller orsakar andra skadliga effekter. Målet om tillförlitlig, etisk och människocentrerad AI kan endast uppnås genom att säkerställa att människor på lämpligt sätt involveras i samband med AI-tillämpningar med hög risk.

Även om de AI-tillämpningar för vilka en specifik rättsordning övervägs i den här vitboken betraktas som tillämpningar med hög risk, kan den lämpliga typen och graden av mänsklig tillsyn variera från ett fall till ett annat. Den ska särskilt bero av hur systemet avses att användas och de effekter användningen kan ha för berörda medborgare och juridiska personer. Den ska inte påverka de lagstadgade rättigheterna enligt den allmänna dataskyddsförordningen i samband med att AI-systemet behandlar personuppgifter. Till exempel kan den mänskliga tillsynen ta sig följande, icke uttömmande, uttryck:

- AI-systemets resultat får inte verkan, om det inte först har granskats och godkänts av en människa (till exempel får en ansökan om socialförsäkringsförmåner avslås enbart av en människa).

- AI-systemets resultat får omedelbar verkan, men ett mänskligt ingripande garanteras i efterhand (till exempel får avslag på en kreditkortsansökan ges av ett AI-system, men en mänsklig granskning måste vara möjlig i efterhand).

- Övervakning av AI-systemet sker när det är i drift, och möjlighet att ingripa i realtid och avaktivera det finns (till exempel en stoppknapp eller ett stoppförfarande i ett förarlöst fordon, om en människa slår fast att bilen inte körs på ett säkert sätt).

- Driftbegränsningar för AI-systemet införs under konstruktionsfasen (till exempel ska en förarlös bil stanna under vissa förhållanden med dålig sikt då sensorerna kan bli mindre tillförlitliga, eller oavsett förhållandena hålla ett visst avstånd till framförvarande fordon).

f) Särskilda krav för biometrisk fjärridentifiering

Insamling och användning av biometriska data [55] för fjärridentifiering [56] , till exempel genom användning av ansiktsigenkänning på allmänna platser, medför särskilda risker för de grundläggande

[55] Biometriska data definieras som ”personuppgifter som erhållits genom en särskild teknisk behandling som rör en fysisk persons fysiska, fysiologiska eller beteendemässiga kännetecken och som möjliggör eller bekräftar identifieringen av den här fysiska personen, såsom ansiktsbilder eller fingeravtrycksuppgifter”. Artikel 3.13 i dataskyddsdirektivet för brottsbekämpning, artikel 4.14 i den allmänna dataskyddsförordningen, och artikel 3.18 i förordning (EU) 2018/1725.

rättigheterna[57]. Konsekvenserna för de grundläggande rättigheterna av AI-system som används för biometrisk fjärridentifiering kan variera avsevärt beroende på syftet, sammanhanget och användningsområdet.

EU:s dataskyddsbestämmelser förbjuder i princip behandling av biometriska data i syfte att entydigt identifiera en fysisk person, utom under särskilda omständigheter[58]. Enligt den allmänna dataskyddsförordningen kan sådan behandling endast ske på ett begränsat antal grunder, varav den viktigaste är hänsyn till ett väsentligt allmänintresse. I så fall måste behandlingen ske på grundval av EU-lagstiftning eller nationell lagstiftning, med beaktande av kraven på proportionalitet, respekt för det väsentliga innehållet i rätten till dataskydd och lämpliga skyddsåtgärder. Enligt dataskyddsdirektivet för brottsbekämpning måste sådan behandling vara absolut nödvändig, i princip ett tillstånd enligt EU-lagstiftning eller nationell lagstiftning, och det måste finnas lämpliga skyddsåtgärder. Eftersom all behandling av biometriska data för att entydigt identifiera en fysisk person skulle hänföra sig till ett undantag från ett förbud som fastställs i EU-lagstiftningen, skulle den omfattas av EU-stadgan om de grundläggande rättigheterna.

Av detta följer att, i enlighet med EU:s befintliga dataskyddsbestämmelser och EU-stadgan om de grundläggande rättigheterna, AI endast får användas för biometrisk fjärridentifiering, om sådan användning är vederbörligen motiverad, proportionerlig och omfattas av lämpliga skyddsåtgärder.

För att ta itu med eventuella samhälleliga problem rörande användningen av AI för sådana ändamål på offentliga platser, och för att undvika en fragmentering på den inre marknaden, kommer kommissionen att inleda en bred europeisk debatt om de eventuella särskilda omständigheter som kan motivera en sådan användning och om gemensamma skyddsåtgärder.

E. ADRESSATER

När det gäller adressaterna för de rättsliga krav som skulle gälla de AI-tillämpningar med hög risk som avses ovan finns det två huvudfrågor att beakta.

För det första är frågan hur skyldigheterna ska fördelas mellan de berörda ekonomiska aktörerna. Många aktörer är involverade i ett AI-systems livscykel. Detta gäller bland annat utvecklaren, spridaren (den person som använder en produkt eller tjänst som är försedd med AI) och potentiellt andra (tillverkaren, distributören eller importören, tjänsteleverantören, den yrkesmässiga eller privata användaren).

Kommissionen anser att varje skyldighet i ett framtida regelverk bör riktas till den eller de aktörer som är bäst lämpade att hantera eventuella risker. Även om till exempel utvecklarna av AI är bäst lämpade att hantera risker med ursprung i utvecklingsfasen, kan deras förmåga att hantera risker under användningsfasen vara mer begränsad. I det fallet bör den relevanta skyldigheten riktas till spridaren.

---

[56] Vid ansiktsigenkänning innebär identifiering att en mall av en persons ansiktsbild jämförs med en mängd andra mallar som lagras i en databas för att ta reda på om hans eller hennes ansiktsbild finns lagrad där. Autentisering (eller verifiering) avser däremot ofta ett-till-ett-matchning. Det gör det möjligt att jämföra två biometriska mallar, som vanligtvis antas tillhöra samma person. Två biometriska mallar jämförs för att fastställa om den person som visas på de två bilderna är samma person. Ett sådant förfarande används till exempel i de automatiska spärrarna vid gränskontrollerna på flygplatser.

[57] Till exempel när det gäller människors värdighet. Rätten till respekt för privatlivet och skyddet av personuppgifter står också i centrum för farhågorna rörande de grundläggande rättigheterna i samband med användning av teknik för ansiktsigenkänning. Det kan också få konsekvenser för icke-diskriminering och rättigheter för särskilda grupper, såsom barn, äldre och personer med funktionsnedsättning. Dessutom får inte yttrande-, förenings- och mötesfriheten undermineras av användningen av tekniken. Se *Facial recognition technology: fundamental rights considerations in the context of law enforcement*, https://fra.europa.eu/en/publication/2019/facial-recognition.

[58] Artikel 9 i den allmänna dataskyddsförordningen, artikel 10 i dataskyddsdirektivet för brottsbekämpning. Se även artikel 10 i förordning (EU) 2018/1725 (tillämplig på EU:s institutioner och organ).

Detta påverkar inte frågan om vilken part som ska vara ansvarig för eventuell uppkommen skada med avseende på ansvar gentemot slutanvändare eller andra parter som lider skada och säkerställande av en effektiv tillgång till rättslig prövning. Enligt EU-lagstiftningen om produktansvar anses tillverkaren bära ansvaret för defekta produkter, utan att det påverkar tillämpningen av nationella lagar som även kan möjliggöra indrivning från andra parter.

För det andra uppkommer frågan om lagstiftningsåtgärdens geografiska tillämpningsområde. Kommissionen anser att det är av största vikt att kraven är tillämpliga på alla relevanta ekonomiska aktörer som tillhandahåller AI-baserade produkter eller tjänster i EU, oavsett om de är etablerade i EU eller inte. I annat fall skulle målen för ovannämnda lagstiftningsåtgärd inte helt kunna uppnås.

F. ÖVERENSSTÄMMELSE OCH EFTERLEVNAD

För att säkerställa att AI är tillförlitligt, säkert och respekterar europeiska värderingar och regler måste de tillämpliga rättsliga kraven uppfyllas i praktiken, och genomdrivas effektivt både av behöriga myndigheter på nationell nivå och EU-nivå och av berörda parter. De behöriga myndigheterna bör vara i stånd att utreda enskilda fall, men också att bedöma effekten på samhället.

Med tanke på den höga risk som vissa AI-tillämpningar utgör för medborgarna och samhället (se avsnitt A ovan) anser kommissionen i det här skedet att en föregående objektiv bedömning av överensstämmelse är nödvändig för att kontrollera och säkerställa att vissa av de ovannämnda obligatoriska kraven för högrisktillämpningar (se avsnitt D ovan) uppfylls. Den föregående bedömningen av överensstämmelse kan omfatta förfaranden för testning, inspektion eller certifiering[59]. Den kan omfatta kontroller av algoritmer och de dataset som används under utvecklingsfasen.

Bedömningarna av överensstämmelse för AI-tillämpningar med hög risk bör ingå i de mekanismer för bedömning av överensstämmelse som redan finns för ett stort antal produkter som släpps ut på EU:s inre marknad. Om inga sådana befintliga mekanismer går att använda, kan liknande mekanismer behöva inrättas som bygger på bästa praxis och eventuella synpunkter från berörda parter och europeiska standardiseringsorgan. Varje sådan ny mekanism bör vara proportionerlig och icke-diskriminerande samt använda transparenta och objektiva kriterier i överensstämmelse med internationella skyldigheter.

Vid utformningen och införandet av ett system som bygger på tidigare bedömningar av överensstämmelse bör särskild hänsyn tas till följande:

- Det kan hända att vissa av de krav som anges ovan inte är lämpliga att kontrollera genom en föregående bedömning av överensstämmelse. Exempelvis lämpar sig kravet på tillhandahållande av information i allmänhet inte väl för kontroll genom en sådan bedömning.

- Särskild hänsyn bör tas till möjligheten att vissa AI-system utvecklas och lär sig av erfarenheter, vilket kan kräva upprepade bedömningar under de berörda AI-systemens livstid.

- Behovet av att kontrollera de data som används för träning och de relevanta metoder, processer och tekniker för programmering och träning som används för att bygga, testa och validera AI-system.

[59] Systemet skulle bygga på förfaranden för bedömning av överensstämmelse i EU, se beslut 768/2008/EG eller förordning (EU) 2019/881 (cybersäkerhetsakten), med beaktande av de särskilda egenskaperna hos AI. Se blåboken om genomförandet av EU:s produktbestämmelser från 2014.

- Om bedömningen av överensstämmelse visar att ett AI-system inte uppfyller kraven för till exempel de data som används för att träna det, måste de konstaterade bristerna åtgärdas, till exempel genom att systemet tränas på nytt i EU på ett sådant sätt att det säkerställs att alla tillämpliga krav uppfylls.

Bedömningen av överensstämmelse skulle vara obligatorisk för alla ekonomiska aktörer som kraven avser, oavsett etableringsort[60]. För att begränsa bördan för små och medelstora företag kan vissa stödstrukturer komma att övervägas, bland annat via de digitala innovationsknutpunkterna. Standarder och särskilda online-verktyg skulle dessutom kunna underlätta uppfyllandet av kraven.

Föregående bedömningar av överensstämmelse bör inte påverka de behöriga nationella myndigheternas övervakning av efterlevnad och efterhandskontroll. Detta gäller för AI-tillämpningar med hög risk, men även för andra AI-tillämpningar som omfattas av rättsliga krav, även om de aktuella tillämpningarnas högriskkaraktär kan utgöra en anledning för de behöriga nationella myndigheterna att ägna särskild uppmärksamhet åt de förstnämnda. Efterhandskontroller bör möjliggöras genom lämplig dokumentation avseende den relevanta AI-tillämpningen (se avsnitt E ovan) och, när så är lämpligt, en möjlighet för tredje parter, till exempel behöriga myndigheter, att testa sådana tillämpningar. Detta kan vara särskilt viktigt när det uppstår risker för de grundläggande rättigheterna, som beror på sammanhanget. Sådan övervakning av efterlevnaden bör ingå i ett system för kontinuerlig marknadsövervakning. Styrningsrelaterade aspekter diskuteras vidare i avsnitt H nedan.

Både för AI-tillämpningar med hög risk och för andra AI-tillämpningar bör dessutom effektiv rättslig prövning säkerställas för parter som påverkas negativt av AI-system. Frågor rörande ansvar diskuteras vidare i rapporten om regelverket för säkerhet och ansvar som åtföljer den här vitboken.

G. FRIVILLIG MÄRKNING AV AI-TILLÄMPNINGAR UTAN HÖG RISK

För AI-tillämpningar som inte klassificeras som ”tillämpning med hög risk” (se avsnitt C ovan) och som därför inte omfattas av de obligatoriska krav som diskuteras ovan (se avsnitten D, E och F) skulle ett alternativ, som ett komplement till gällande lagstiftning, vara att införa ett frivilligt märkningssystem.

Enligt systemet skulle berörda ekonomiska aktörer som inte omfattas av de obligatoriska kraven kunna besluta att på frivillig basis omfattas av antingen de här kraven eller en särskild uppsättning liknande krav som särskilt fastställts för det frivilliga systemet. De berörda ekonomiska aktörerna skulle sedan tilldelas en kvalitetsmärkning för sina AI-tillämpningar.

Den frivilliga märkningen skulle göra det möjligt för de berörda ekonomiska aktörerna att signalera att deras AI-baserade produkter och tjänster är tillförlitliga. Det skulle göra det möjligt för användarna att enkelt se att de berörda produkterna och tjänsterna överensstämmer med vissa objektiva och standardiserade EU-övergripande riktmärken som går längre än de normalt tillämpliga rättsliga skyldigheterna. Detta skulle bidra till att öka användarnas förtroende för AI-system och främja den övergripande spridningen av tekniken.

Det här alternativet skulle innebära att det skapas ett nytt rättsligt instrument som fastställer ramen för frivillig märkning för utvecklare och/eller spridare av AI-system som inte betraktas som system med

[60] När det gäller den relevanta styrningsstrukturen, inklusive de organ som utsetts att utföra bedömningarna av överensstämmelse, hänvisas till avsnitt H nedan.

hög risk. Även om deltagande i märkningssystemet skulle vara frivilligt, skulle kraven vara bindande så snart utvecklaren eller spridaren valt att använda märket. Det behöver säkerställas att alla krav uppfylls genom en kombinationen av förhands- och efterhandskontroller.

H. STYRNING

En europeisk styrningsstruktur för AI i form av en ram för samarbete mellan nationella behöriga myndigheter är nödvändig för att undvika fragmentering av ansvarsområdena, öka kapaciteten i medlemsstaterna och säkerställa att Europa utrustar sig successivt med den kapacitet som krävs för testning och certifiering av AI-baserade produkter och tjänster. I det här sammanhanget skulle det vara fördelaktigt att stödja de behöriga nationella myndigheterna så att de kan fullgöra sitt uppdrag där AI används.

En europeisk styrningsstruktur skulle kunna ha en rad olika uppgifter, som ett forum för regelbundet utbyte av information och bästa praxis, med kartläggning av nya trender, rådgivning om standardisering samt om certifiering. Den bör också spela en nyckelroll när det gäller att underlätta genomförandet av den rättsliga ramen, till exempel genom att utfärda riktlinjer, avge yttranden och göra expertutlåtanden. Den bör därför förlita sig på ett nätverk av nationella myndigheter, sektorsspecifika nätverk och tillsynsmyndigheter på nationell nivå och EU-nivå. Dessutom skulle en expertkommitté kunna bistå kommissionen.

Styrningsstrukturen bör garantera ett så högt deltagande av berörda parter som möjligt. Berörda parter – konsumentorganisationer och arbetsmarknadens parter, företag, forskare och organisationer i det civila samhället – bör rådfrågas om genomförandet och vidareutvecklingen av ramen.

Med tanke på att det redan finns strukturer inom till exempel finansiering, läkemedel, luftfart, medicintekniska produkter, konsumentskydd och dataskydd bör den föreslagna styrningsstrukturen inte överlappa de befintliga funktionerna. Den bör i stället upprätta nära förbindelser med andra behöriga myndigheter på EU-nivå och nationell nivå inom de olika sektorerna för att komplettera befintlig expertis och hjälpa befintliga myndigheter att övervaka och utöva tillsyn över den verksamhet som bedrivs av ekonomiska aktörer och som involverar AI-system och AI-baserade produkter och tjänster.

Slutligen skulle, om det här alternativet väljs, genomförandet av bedömningar av överensstämmelse kunna anförtros anmälda organ som utsetts av medlemsstaterna. Testcenter bör möjliggöra en oberoende revision och bedömning av AI-system i enlighet med de krav som anges ovan. En oberoende bedömning kommer att öka förtroendet och garantera objektivitet. Den kan också underlätta arbetet för de relevanta behöriga myndigheterna.

EU har utmärkta center för testning och bedömning och bör utveckla sin kapacitet även på AI-området. Ekonomiska aktörer som är etablerade i tredjeländer och som vill ta sig in på den inre marknaden skulle antingen kunna anlita särskilt utsedda organ som är etablerade i EU eller, med förbehåll för avtal om ömsesidigt erkännande med tredjeländer, anlita organ från tredjeländer för att genomföra sådan bedömning.

Styrningsstrukturen för AI och de eventuella bedömningar av överensstämmelse som det är fråga om här skulle inte påverka de befogenheter och det ansvar som de relevanta behöriga myndigheterna skulle ha enligt gällande EU-lagstiftning i specifika sektorer eller specifika frågor (finans, läkemedel, luftfart, medicintekniska produkter, konsumentskydd, dataskydd osv.).

6. SLUTSATS

AI är en strategisk teknik som erbjuder många fördelar för medborgarna och företagen och för samhället som helhet, förutsatt att den är människocentrerad, etisk, hållbar och respekterar grundläggande rättigheter och värden. AI erbjuder viktiga effektivitets- och produktivitetsvinster som kan stärka den europeiska industrins konkurrenskraft och främja medborgarnas välbefinnande. Det kan också bidra till att finna lösningar på några av de mest trängande samhällsutmaningarna, såsom kampen mot klimatförändringar och miljöförstöring, utmaningarna i samband med hållbarhet och demografiska förändringar, och skyddet av våra demokratier och, när så är nödvändigt och proportionerligt, brottsbekämpning.

För att kunna utnyttja de möjligheter som AI erbjuder måste Europiska unionen utveckla och förstärka den nödvändiga industriella och tekniska kapaciteten. Såsom anges i den åtföljande europeiska strategin för data krävs det också åtgärder som gör det möjligt för EU att bli ett globalt nav för data.

Den europeiska AI-strategin syftar till att främja Europas innovationskapacitet på AI-området och samtidigt stödja utvecklingen och användningen av etisk och tillförlitlig AI i hela EU:s ekonomi. AI ska vara till stöd för människorna och en positiv kraft i samhället.

Med den här vitboken och den åtföljande rapporten om ramen för säkerhet och ansvar lanserar kommissionen ett brett samråd med medlemsstaterna, det civila samhället, näringslivet och den akademiska världen om konkreta förslag till en europeisk AI-strategi. De omfattar både politiska medel för att främja investeringar i forskning och innovation, förbättra kompetensutveckling, stödja små och medelstora företags användning av AI, och förslag till viktiga delar i ett framtida regelverk. Samrådet kommer att möjliggöra en omfattande dialog med alla berörda parter som kommer att bilda underlag för kommissionens nästa steg.

> Kommissionen inbjuder alla att komma med synpunkter på förslagen i vitboken genom ett öppet offentligt samråd som finns tillgängligt på: https://ec.europa.eu/info/consultations_en. Till och med den 19 maj 2020 går det att lämna synpunkter inom ramen för samrådet.
>
> Det är normal praxis för kommissionen att offentliggöra synpunkter som inkommit i samband med ett offentligt samråd. Det är dock möjligt att begära att synpunkterna eller delar av dessa förblir konfidentiella. Om så är fallet, ange tydligt på första sidan av er inlaga med synpunkter att ni begär att den inte ska offentliggöras. I det fallet bör också en icke-konfidentiell version av synpunkterna översändas till kommissionen för offentliggörande.