Europeiska kommissionens förslag till förordning om harmoniserade regler för artificiell intelligens, Proposal for a Regulation laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts (COM(2021) 206)

Remissvar av WASP-HS

Kontakt: Professor Virginia Dignum, Umeå universitet – Programdirektör WASP-HS, virginia@cs.umu.se

Det är inte någon nyhet att teknik, metoder och applikationer med artificiell intelligens (AI) påverkar samhället och individer. Potentialen med AI är tyvärr inte utan risker och utmaningar, vilket många exempel på missbruk, lagöverträdelser och negativ påverkan på mänskliga rättigheter, samhälleliga och etiska värden och miljön, har visat. Liksom tidigare, och trots stora framsteg inom exempelvis automatisering, visualisering, planering och spel, så lovar AI för mycket och levererar för lite – precis som andra tekniker till vilka folk har haft stora förhoppningar. Med kommissionens förslag för en förordning för harmoniserade regler om artificiell intelligens (”Artificial Intelligence Act”, AIA) visar EU att mänskliga rättigheter och europeiska värderingar ska vara kärnan i utvecklingen och användningen av AI, som först och främst ska appliceras på ett människo-centrerat sätt och som en positiv kraft i samhället med det slutgiltiga målet att öka mänskligt välbefinnande.

The Wallenberg Program on AI, Autonomous Systems and Software Program – Humanities and Society (WASP-HS) välkomnar att AIA inte bara ser till riskerna som är kopplade till AI utan också höjer ribban väsentligt vad gäller kvalitet, prestanda och pålitlighet för AI som EU är beredd att tillåta. Listan över förbjudna AI-tillämpningar, vid sidan av kraven på AI med höga och medelhöga risker, samt åtgärderna för att främja pålitlig AI-innovation, däribland modellen med frivillig efterlevnad, kommer att vara en del i förbättringen av kvalitet, prestanda och pålitlighet för AI. Detta kommer att bidra till att AI kan leva upp till de förväntningar som finns, så samhället kan dra nytta av AI med minimerade risker.

Vår rekommendation fokuserar därför inte på målen och grundtanken med AIA utan helt och hållet på genomförandet. AIA samlar delar av lagstiftning gällande mänskliga rättigheter med en risk-baserad inställning och ett områdesbaserat fokus, med en väldigt bred definition av AI-teknik. Denna kombination av aspekter, tillsammans med av AIA och existerande förordningar kan potentiellt leda till en alltför komplex struktur av regelverk som kommer att vara svår att till fullo behandla, kontrollera förenligheten av och kan skapa kryphål i implementeringen.

WASP-HS skulle vilja att den svenska regeringen tar extra hänsyn till följande aspekter:

1. Definitionen av AI som används i AIA är både för bred och för smal. Vi anser att det är extremt svårt att identifiera vilka typer av programvara som inte täcks av definitionen. Samtidigt är AI, i dess effekt på samhället, mer än bara en teknik och skulle istället kunna ses som ett socio-tekniskt ekosystem. AIA borde i grunden satsa på att säkerställa tillit i utvecklingen och användandet av AI, genom att klargöra ansvaret som de aktörer har som utvecklar, tillhandahåller eller nyttjar AI. Förtroendet för AI måste komma från förtroende för det socio-tekniska systemets implementering av AI. Det blir mer och mer tydligt att problem som fördomar, diskriminering och säkerhet inom AI-system inte kan lösas endast via tekniken utan det finns behov för institutionella åtgärder och organisatoriska åtaganden. Å enda sidan borde det inte, från ett europeiskt innevånarperspektiv, spela någon roll om denna teknik används för göra automatiska bedömningar av dem, utan det viktiga borde vara att det finns tillräckliga skyddsåtgärder och kontroller för att garantera att bedömningarna är pålitliga och att de ansvariga kan stå till svars för resultatet. Å andra sidan kan en för bred definition av AI-system leda till (alltför) mycket arbete för europeiska företag och offentliga organisationer med kontroller och åter-certifiering av redan existerande programvara som faller under denna breda definition.
2. Att betrakta och behandla AI som inbäddat i ett socio-tekniskt ekosystem och betona de vetenskapliga aspekterna av AI-utveckling istället för att begränsa AI till en teknisk disciplin skulle bidra till ett mer mångfaldigt och inkluderande deltagande, där en bättre balans skulle kunna uppnås vad gäller representation, bakgrund och forskningsområde bland de som utvecklar, tillhandahåller och använder AI. Ett bredare perspektiv på området AI skulle också bidra till en bättre inställning gällande data governance och validiteten av tillvägagångsätt.
3. Istället för att begränsa möjligheter borde AIA utformas som en språngbräda för hållbar AI-innovation. De brister som kommer med de existerande data-drivna AI-metoderna blir alltmer tydliga, och är delvis ett resultat av de stokastiska egenskaper de dataanalysmetoder som används har i kombination med storleken på beräkningarna. Sådana metoder fungerar väl när det gäller precision men gör mycket sämre ifrån sig när det handlar om transparens och förklaringar. Ett starkt engagemang och stort stöd för forskning och innovation i alternativa AI-metoder, som kan kombinera precision med transparens och personskydd, samt multi-disciplinära insatser för att utveckla och utvärdera AI;s samhälleliga och etiska konsekvenser, måste vara en central del av operationaliseringen av AIA. Istället för att bara följa med i de rådande trenderna med data-driven AI har europeisk forskning inom AI möjlighet och kapacitet att utöka och förbättra nuvarande synsätt till AI 2.0 – ett verkligt människocentrerat AI. Denna kapacitet måste näras och stödjas med stora investeringar från EU och på nationell nivå och i samarbeten mellan privata företag och det offentliga.

4. Upprätthållandet av förpliktelser och stöttning vad gäller ansvar för AI och dess effekter är ytterst viktigt. Den viktigaste frågan för oss gäller operationaliseringen av dessa principer. Om det ibland är svårt att identifiera risker och hot från en praktisk synvinkel kan det också vara svårt för i synnerhet små och medelstora företag att visa på hur de efterlever kraven när det finns så många undantag och inbördes förhållanden att ta hänsyn till. AIA innehåller ett flertal bedömnings- och kontrolleringsförpliktelser. En kombination av den föreslagna riskbaserade inställningen och en kontextmedveten implementering bidrar till att identifiera möjligheter istället för att fokusera på risker. Utvecklandet och tillhandahållandet av tjänster och tekniska verktyg för att hjälpa organisationer – i synnerhet små och medelstora företag, nystartade företag och mikroorganisationer – att uppfylla de krav som finns är ett potentiellt nytt område att ombesörja och främja.
5. AI verkar redan nu i en värld som inte är laglös. Eventuella överlappningar, luckor och bristande överenstämmelser mellan AIA och existerande lagar (GDPR, Europeiska unionens stadga om de grundläggande rättigheterna, men också nationell arbetsrätt, förvaltningsrätt och så vidare) behöver utvärderas och övervakas kontinuerligt.
6. Vi är positiva till en så kallad sandbox-inställning vid utveckling och användning av AI-system i högriskmiljöer. Detta måste bestå av medverkan från flertalet intressenter och erfarenheter och utvärderingar från flera olika discipliner, snarare än att endast fokusera på de tekniska/datatekniska aspekterna på lösningar.
7. Med tanke på hur viktiga infrastrukturer för kommunikation är och hur känsliga de är för AI-drivna attacker och missbruk känner vi oss bekymrade över att denna infrastruktur inte finns med på listan över livsviktiga infrastrukturer i bilaga III. Dessutom, med tanke på den potentiellt enorma påverkan nyttjande av AI i utbildningssammanhang kan ha på formandet av kommande generationer, borde användandet av oövervakade autonoma system för undervisning också räknas som högrisk.
8. Slutligen måste omfattande utbildningsinsatser vara en del av implementeringen av AIA. Alla europeiska innevånare måste ha en grundläggande förståelse för tekniken som ligger bakom de automatiserade beslutsfattande systemen och ha möjlighet att lita tillräckligt på sig själva för att kunna avgöra om besluten verkar resonabla eller inte, och att kunna ifrågasätta och till och med misstro systemet. Tilltro till systemets beslut är en sak, men tilltro till det egna användandet av systemet är en annan. Detta är en fundamental förutsättning för en tillförlitlig utveckling och användning av AI.