Mellan juridisk fixering och flexibilitet i den europeiska AI-förordningen

Larsson, Stefan

Published in:
Advokaten. Tidskrift för Sveriges advokatsamfund

2024

Document Version:
Förlagets slutgiltiga version

Mellan juridisk fixering och flexibilitet i den

Det är mycket rörelse på AI-fronten nu – både rättsligt, tekniskt och politiskt. Stefan Larsson, jurist och docent i teknik och social förändring vid Lunds universitet, placerar in **EU-lagstiftarens AI-förordning i en både global och svensk kontext**. Han beskriver processen bakom lagstiftningen, de etiska och principiella grunderna, och hur EU-lagstiftaren hanterar AI-innovationernas explosionsartade rörlighet med flexibla element i AI-förordningen, som på samma gång kan skapa dynamisk styrning och rättslig oförutsägbarhet.[1]

Introduktion: AI i fokus

Fredagen den 2 februari i år godkände representanterna för de 27 länderna i EU enhälligt världens första heltäckande regelbok för artificiell intelligens, vilket knöt ihop säcken för den politiska överenskommelse som nåddes i december. Med detta nåddes därmed en slutpunkt för långa och delvis tämligen högröstade förhandlingar i den så kallade trilogen, där starka motsättningar fanns kring bland annat hur undantagen för förbud mot biometrisk realtidsidentifiering i publika miljöer skulle formuleras och om de största, mer allmänna, AI-modellerna skulle regleras eller ej. Det betyder att EU:s AI-förordning allra troligast träder ikraft inom kort och börjar tillämpas i sina olika delar över de kommande tre åren.

För svenskt vidkommande har vi en nytillsatt AI-kommission, som är en statlig kommitté med uppdrag att ge förslag på hur Sverige både ska underlätta innovation och effektivisera offentlig förvaltning och samtidigt identifiera hur Sverige kan agera för att främja en konkurrenskraftig och säker AI inom EU och globalt. Det är stora ord, och det blir intressant att följa kommitténs arbete. Internationellt pågår ett race för teknikföretagen att ligga i framkant av AI-utvecklingen inom så kallade generativa AI-modeller, samtidigt som en rad regleringsbehov uttrycks i allt från rapporter från OECD och FN till riktlinjer från G7 och president Bidens presidentorder från oktober om säker och tillförlitlig AI. Farhågorna handlar om risker för missinformation, det vill säga medvetna försök att till exempel manipulera och undergräva nyhetsflöden och demokratiska processer, men även diskriminerande utfall kopplade till brist på transparens, och hur olika rättigheter till träningsdata bör hanteras. Kort sagt, AI-frågorna är i fokus för både innovatörer och reglerare. Både globalt, i Sverige och övriga Europa.

Stefan Larsson är jurist och docent i teknik och social förändring vid institutionen för teknik och samhälle på LTH, Lunds universitet. Han leder där en forskargrupp inriktad mot AI och samhälle. Han är både filosofie doktor i rättssociologi och teknologie doktor i fysisk planering, och forskargruppens forskning är inriktad mot frågor om AI-transparens, tillit och olika sorters styrning och normer i användningen av AI och automation.

Den europeiska AI-förordningen

EU-lagstiftarens AI-förordning, även kallad rättsakten om artificiell intelligens eller AIA, helt kort, har föregåtts av en påtagligt intensiv debatt både utom och inom trilogen mellan EU:s lagstiftande institutioner under hösten 2023. Eftersom ingen offentlig och slutgiltig version ännu finns publicerad vid tillfället för denna artikels färdigställande (7/2), utgår jag i denna analys från den version som antogs i Coreper I, det vill säga rådets kommitté med medlemsstaternas ständiga representanter, fredagen den 2 februari 2024. Mycket kan dock först sägas om utmaningarna överlag att reglera AI, och det så kallade taktproblem det innebär, det vill säga att lagstiftaren har att försöka fånga det rörliga mål som AI-utvecklingen utgör.

I denna artikel beskriver jag dels på vilket sätt AI-fältet är rörligt – tänk stora språkmodeller och ”general purpose AI” – och med vilka flexibla, ”framtidssäkra” (i EU-lingo) mekanismer EU-lagstiftaren hanterar marknadens rörlighet. För att kunna göra det behöver jag dels placera in den nya EU-regleringen i en global kontext, dels göra en kort återblick till de etiska och principiella riktlinjer som i mycket angett tonen och etablerat några av de grundläggande idéerna för AI-rättsakten.

Utmaningarna med att reglera artificiell intelligens är flera. *För det första* är själva begreppet högst dynamiskt. Det vill säga vad som räknas som AI har varierat över tid, alltsedan begreppet etablerades som ett forskningsfält under 1950-talet. Eftersom EU-kommissionen under 2021 bedömde att de samhällsutmaningar – inte bara möjligheter – som medföljer AI bäst borde regleras genom att reglera AI som sådant, snarare än de beteenden som inte är

FOTO: SARA ARNALD

europeiska AI-förordningen

1 Artikeln bygger på rapporten *Reglering av AI: för lite för sent, eller för mycket för tidigt? En rapport om generativ AI och den europeiska AI-förordningen*, som har skrivits av Stefan Larsson. Den publicerades den 15 november 2023 av Tillväxtanalys, myndigheten för tillväxtpolitiska utvärderingar och analyser. Analysen har också utförts inom forskningsprojektet The Automated Administration: Governance of Automated Decision-Making in the Public Sector som leds av Stefan Larsson. Projektet finansieras av forskningsprogrammet Future Challenges in the Nordics (futurenordics.org). Tack riktas särskilt till Jockum Hildén, pol. dr och associate på Mannheimer Swartling, för läsning av underliggande rapport, och idémässiga bidrag.

önskvärda, hamnade AI-förordningen i en terminologisk kamp om att definiera AI.

För det andra är fältet påtagligt intensivt i sin utveckling, vilket nog inte undgått någon i termer av den utveckling med stora språkmodeller och så kallad generativ AI som OpenAI och ChatGPT i mycket fått klä skott för. Detta gör att regleraren har svårt att med exakthet fånga på vilket sätt man bäst med lagstiftning bör styra för att på samma gång möjliggöra innovationerna och om inte omöjliggöra, så åtminstone dämpa, risker och missbruk av tekniken. Denna temporala skillnad, enklast uttryckt som ”snabb innovation, långsam lagstiftning”, kan kallas för ett taktproblem, där lösningarna brukar stavas olika grad av flexibilitet, som ”anticipatory regulation”, experimentella upplägg med regulatoriska sandlådor eller den dynamik som kan tillåtas i mjuk lagstiftning och uppförandekoder. Denna flexibilitet kan tyckas både aptitlig och gångbar på områden som utmärks av stor rörlighet och där de normativa svaren inte tycks färdiga, men baksidan på samma mynt att balansera mot är bristande rättslig förutsägbarhet och därmed ytterst rättsosäkerhet, vilket vi återkommer till nedan.

För det tredje, och avslutningsvis, är AI-fältet en del i en global kapprustning med både ekonomiska och säkerhetspolitiska förtecken. Här är inte minst USA, Kina och EU de mest betydande ekonomier som positionerar sig på olika sätt gentemot varandra. Det finns inte utrymme att fördjupa den analysen här, mer än att konstatera att det påverkar EU:s strategier och rättsliga utveckling. Låt oss först fokusera taktproblemet.

Snabb teknikmedierad förändring och behovet av juridisk förutsägbarhet

Taktproblematiken mellan AI-innovation och EU-lagstiftning ställdes på sin spets under 2023. Om man jämför AI-förordningens tillblivelse med AI-utvecklingen kan man först konstatera att EU-lagstiftarens tre förslag kom i april 2021 (kommissionen), december 2022 (rådet) och juni 2023 (parlamentet) för att förhandlas i trilog under hösten 2023. Ett första gemensamt förslag nåddes den 9 december 2023 efter en beryktad sista sittning på 33 timmar, med avbrott endast för sömn.

Om de första två versionerna i huvudsak var inriktade mot att fördela olika AI-användningsområden utifrån en riskskala, möttes parlamentet under våren 2023 av utvecklingen inom vad som kommit att kallas generativ AI. Det vill säga större AI-modellers mer allmänna förmågor att skapa nya texter, bilder, programkod med mera baserat på mänskliga instruktioner, så kallade prompter. Den kapacitetsökningen, kombinerat med användarvänligheten och den tillgänglighet som en enkel webbaccess skapade, gjorde att den kanske mest kända, ChatGPT, redan två månader efter lanseringen i november 2022 nådde över 100 miljoner unika besökare. Det är enligt vissa bedömare den snabbaste tillväxten i användning av någon enskild teknologi genom alla tider. Detta faktum landade, så att säga, på EU-parlamentets bord.

Farhågorna handlar om risker för missinformation, det vill säga medvetna försök att till exempel manipulera och undergräva nyhetsflöden och demokratiska processer, men även diskriminerande utfall kopplade till brist på transparens, och hur olika rättigheter till träningsdata bör hanteras.

ChatGPT och de underliggande GPT-modellerna (för närvarande GPT4), utvecklade av OpenAI med stöd från Microsoft, är bara toppen på ett isberg av aktörer som utvecklar avancerade AI-modeller. Det innehåller språkmodeller som Metas LLaMa-varianter, Googles Bard (vilket nyligen döptes om och inkluderades i Gemini), Anthropics Claude-varianter, HuggingFaces BLOOM eller kinesiska varianter som Huaweis PanGu-Σ eller Baidus Ernie 4.0 (som bland annat ger chattbotten Ernie Bot sin funk-

- World Economic Forum bjöd i april 2023 in en rad experter som bistod i att ta fram en policyrapport som publicerades efterföljande juni med titeln *The Presidio Recommendations on Responsible Generative AI*. Här argumenteras bland annat för behovet av AI-litteracitet allmänt och mer expertis hos beslutsfattare, med betoning på ansvarsfull utveckling och distribution av generativ AI överlag, men även öppen innovation och internationellt samarbete samt samhällsnytta.
- OECD publicerade *Initial policy considerations for generative artificial intelligence* i september 2023. Här pekas bland annat på hur generativ AI kan ge stora värden, med revolutionerande konsekvenser för industrier, men att det även medför risker. Dessa pekas på i termer av missinformation, reproduktion av bias, det vill säga diskriminerande strukturer, immaterialrättsliga utmaningar och några andra policyrelevanta frågor.
- Teknikprofessionsorganisationen ACM:s teknologipolicyråd publicerade i juni 2023 *Principles for the development, deployment, and use of generative AI technologies*, där de bland annat pekar på behovet av reglering och införandet av skyddsmekanismer, med human-in-the-loop, utmaningar med frågor kopplat till immaterialrättsligt ägande, dataskydd och behov av möjligheter och mekanismer för att korrigera felaktigheter.
- President Biden utfärdade en presidentorder om säker och tillförlitlig artificiell intelligens den 31 oktober 2023, som bland annat ställer krav på att de mäktigaste utvecklarbolagen delar kritisk information med federala myndigheter, på behovet av att utveckla standarder, och på märkning av AI-genererat innehåll.
- G7 publicerade AI-riktlinjer för en "Hiroshimaprocess" för avancerade AI-system samt en uppförandekod för utvecklarorganisationer den 30 oktober 2023, med betoning på transparensåtgärder som informationsdelning, men även frågor om autentisering och immaterialrättigheter i träningsdata.
- De länder som närvarade vid Storbritanniens AI safety summit skrev den 1 november 2023 under *The Bletchley Declaration* som bland annat efterlyser internationellt samarbete för att hantera AI-risker och särskilt pekar ut flerändamåls-AI och AI-grundmodeller.
- Förenta Nationerna tillsatte den 26 oktober 2023 ett rådgivande organ för artificiell intelligens, vars första uppgift blev att ta fram och publicera *Interim Report: Governing AI for Humanity*, som fokuserar mycket på vikten av att inkludera alla medborgare, även på en global skala, med betoning på internationell rätt och datahanteringens centrala betydelse.

tionalitet). Liknande underliggande modellträning ligger även till grund för AI-system som kan generera bilder genom prompter, som DALL·E 3, Midjourney och Stable Diffusion. Även ljud kan instrueras eller "promptas". Exempelvis VALL-E-modellen, som Microsoft publicerade i januari 2023, kräver bara 3 sekunder från en ljudinspelning för att realistiskt kunna generera nya uttryck i samma stil. Både nyttor och problem bör ses i ljuset av den distribution som når miljarder människor snarare än miljoner genom integrering i sökmotorer, mjukvarusviter och spridning i sociala medier i övrigt.

Det betyder att stora belopp investeras i modellutvecklingen och dess underliggande infrastruktur. Man kan föreställa sig både mindre resurskrävande men mer kapabla modeller, vilket är relevant för hur EU-lagstiftaren valt att reglera denna typ av modeller. Vi återkommer till det nedan. Man kan också förvänta sig olika varianter av integrering i befintliga tjänster som kan hjälpa till att skriva, översätta talat språk i realtid, skapa fejkade men realistiska videos, utveckla bildverktyg med mera, vilket är relevant för hur man ser på ansvarsfördelning och olika aktörers behov av transparens i denna typ av ekosystem. Multimodala modeller för mer allmänt bruk, det vill säga sådana som kan växla mellan eller kombinera ljud, bild och video lite mer dynamiskt, är att vänta under 2024, exempelvis i Googles Geminivarianter.

Allt detta tog ett starkt hopp framåt under själva lagstiftningsprocessen med AI-förordningen i EU. Även om rådets version av AI-förordning från december 2022 hade nämnt behovet av att reglera AI-modeller av mer allmän karaktär ("general purpose AI") så kände sig parlamentet nödgat att fånga riskerna med dessa mer generella AI-modellers kapacitet på ett tydligare sätt. Svaret i parlamentets version blev att föreslå en rad transparens- och dokumentationskrav på vad parlamentet kallade "grundmodeller" ("foundation models") och generativ AI.

Kartläggningar av den här typen av principiella dokument och policyer visar på hur såväl företag och stater som civilsamhället formerade AI-riktlinjer. Etik, tycks det, blev ett sätt att utveckla styrning av AI. Så även inom EU.

Just denna del av regleringen kom också att bli en av de hetaste förhandlingsfrågorna under höstens trilog, där framförallt Frankrike och Tyskland, med stöd från Italien, rapporterades vilja undvika den typen av generell reglering (rapporteringen hänvisade till tryck från industriella företrädare i Tyskland och Frankrike som utvecklar just flerändamåls-AI). Slutsatsen i förhandlingarna – det fanns naturligtvis en rad andra frågor på bordet som rönte stort intresse, inte minst frågan om biometrisk fjärridentifiering för brottsbekämpning – blev likväl en rad krav vid utveckling av AI-modeller av allmän karaktär.

Någon officiell svensk översättning finns inte vid skrivande av denna text, men man kan konstatera att parlamentets terminologi kring "grundmodeller" inte har inkluderats i den version som godkändes i Coreper I den 2 februari, 2024. Istället har begreppen "general purpose AI models" eller "GPAI models", "GPAI systems" och den helt nya kategorin "GPAI models with systemic risk" introducerats. Sistnämnda medför en nivåindelning i regleringen av dessa AI-modeller av allmännare slag, där de allra mest kraftfulla enligt lagstiftningen medför risker av "systemisk" karaktär, vilket medför fler krav på hantering. Systemisk risk medför enligt skälen i AI-förordningen "negativa effekter på folkhälsa, säkerhet, allmän säkerhet, grundläggande rättigheter eller samhället som helhet, som kan spridas i stor skala över hela värdekedjan" (författarens översättning). Gränsdragningen sker genom en bedömning av hur mycket beräkningskapacitet (mätt i så kallade FLOPS, floating point operations per second) som använts vid träningen. EU-kommissionen ges dock i rättsakten mandat att ändra på både nivåer och hur bedömningen för "systemisk risk" ska gå till, vilket är en intressant flexibilitet jag återkommer till i analysen nedan.

Innovationens etik och riskbild

Var kommer då idéerna bakom AI-förordningen ifrån? Det tycks finnas två vågor av etiska riktlinjer och "soft policy", det vill säga mjuk AI-styrning, dels från åren innan kommissionens förslag på AI-förordning (2021), dels under 2023, som en följd av utvecklingen inom generativ AI, vilket redovisats ovan.

Åren 2016 till omkring 2020 kan beskrivas som väldigt intensiva när det gäller framtagandet av etiska riktlinjer och principiella ställningstaganden gällande AI-utveckling och AI-tillämpning. En medvetenhet om inte bara möjligheter utan även risker för till exempel diskriminerande praktik i allt från AI-system för ansiktsigenkänning till sökmotorer,

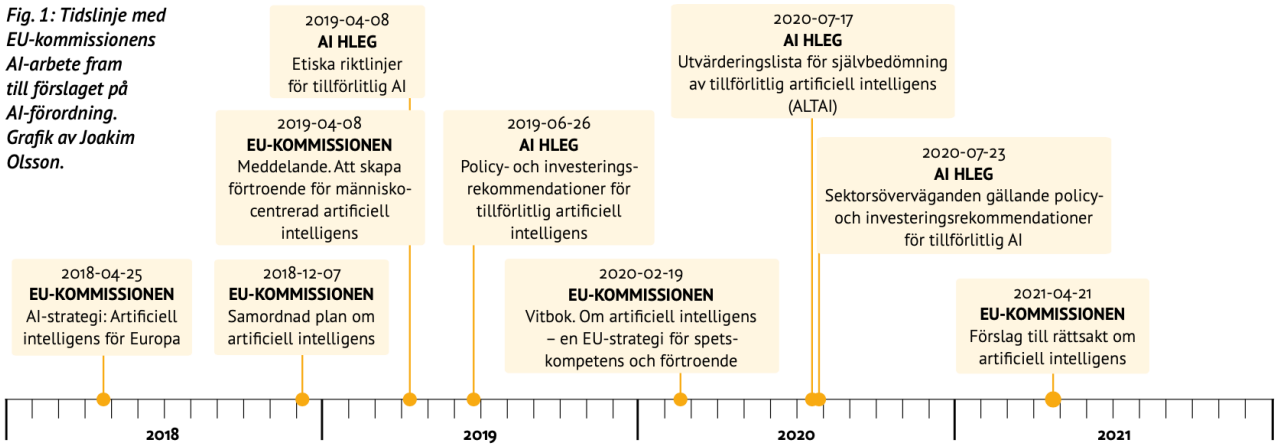

Fig. 1: Tidslinje med EU-kommissionens AI-arbete fram till förslaget på AI-förordning. Grafik av Joakim Olsson.

myndighetsbeslut och HR-system spred sig. För sistnämnda kan man med fördel läsa Diskrimineringsombudsmannens rapport till regeringen från november 2023, *AI och risker för diskriminering i arbetslivet* (2023:6).

Kartläggningar av den här typen av principiella dokument och policyer visar på hur såväl företag och stater som civilsamhället formerade AI-riktlinjer. Etik, tycks det, blev ett sätt att utveckla styrning av AI. Så även inom EU. I denna typ av riktlinjer har betydelsen av transparens i AI-utveckling och AI-implementering betonats, med särskilda utmaningar gällande ansvarsfördelning i automatiserade system, baserat på modeller som ofta beskrivs som ”black box”, det vill säga med inbyggda svårigheter att härleda var träningsdatamängden kommer ifrån och hur den är konstituerad – vilket är relevant för både bedömning av diskriminerande utfall, såväl som till exempel efterlevnad av dataskydd och upphovsrätt. Bristen på transparens kan även gälla vilka kompetenser som ingår i träningsfasen och hur upplägg med algoritmer, träningsdata och inbyggda målsättningar för AI-modellerna ser ut (gällande de mångfacetterade intressen som möts och i någon mån måste balanseras, se min och professor Fredrik Heintz artikel ”Transparency in Artificial Intelligence” i *Internet Policy Review* 2020. Gällande spänningsfältet mellan teknisk förklarbarhet och mänsklig litteracitet, se gärna mitt och mina kollegors kapitel om ”Four Facets of AI Transparency” från 2023.)

Kommissionen inrättande en AI-expertgrupp (HLEG) som under 2019 och 2020 publicerade flera rapporter (se fig. 1), där *Etiska riktlinjer för tillförlitlig AI* tycks ha haft särskilt betydande roll för AI-styrningsområdet. Den hänvisas också till i AI-förordningens skäl och i något allmännare ordalag i artikel 69, som behandlar de uppförandekoder som ska tas fram.

Under 2023 blev den globala debatten kring behovet av reglering av AI särskilt tongivande, som en följd av hur generativ AI, ChatGPT och de olika bakomliggande GPT-modellerna blev särskilt tydliga symboler för en snabb utveckling med både nyttor och risker. Det kom att uttrycka en ny våg av policyer, som också kom att påverka EU-lagstiftaren gällande AI-förordningen, som nämnt ovan.

I forskarvärlden syntes en uppdelning mellan dem som såg mer spekulativa existentiella risker och dem som såg mer handfasta men oerhört allvarliga problem med bias och maktförskjutningar. En av de mest omtalade händelserna kopplad till de förstnämnda var det så kallade moratoriebrevet, som publicerades den 22 mars 2023, med över 31 000 signaturer. Undertecknarna krävde en paus på åtminstone sex månader för träning av AI-system som är mer kraftfulla än OpenAI:s GPT-4-modell. Strax därefter, i april 2023, argumenterade samhällsvetenskapliga forskare kopplade till forskningsinstitutet AI Now för att den europeiska AI-förordningen behöver adressera ”general purpose AI” ur synvinkeln att risker för skadlig och diskriminerande bias förstärks av AI-modellerna. Detta, menade forskarna, riskerar att leda till att utvecklare och leverantörer av AI-modeller undviker ansvar för de förödande konsekvenser AI kan ha på individnivå.

”Alltsedan kommissionen publicerade sin AI-strategi under 2018 har arbetet med ett slags dubbel hantering av AI-frågorna pågått, av å ena sidan att försöka stimulera innovation och å andra sidan att begränsa risker.

Utvecklingen inom generativ AI har också föranlett en rad aktörer att ta fram policyer och rekommendationer under 2023 för hur grundmodeller eller flerändamåls-AI bör hanteras, se faktaruta överst på s. 40. Dessa policyer understryker att vi befinner oss i en formativ period gällande styrning och reglering av generativ AI.

Från etik till juridik: Den europeiska AI-förordningen

Alltsedan kommissionen publicerade sin AI-strategi under 2018 har arbetet med ett slags dubbel hantering av AI-frågorna pågått, av å ena sidan att försöka stimulera innovation och å andra sidan att begränsa risker. För en tidslinje, se fig. 1. Detta är tydligt även i den vitbok om AI som kommissionen publicerade 2020.

I vitboken och de etiska riktlinjerna syns flera av de idéer om risknivåer och transparens som fångats i den AI-förordning som fastställer skyldigheter för leverantörer och användare beroende på risknivån förknippad med olika sorters AI-system. Förordningen gäller leverantörer som släpper ut AI-system på marknaden och tar dem i bruk i unionen, oavsett om dessa leverantörer är etablerade inom unionen eller i ett tredjeland. De olika risknivåerna beskrivs som *oacceptabel risk*, vilket motsvaras av förbud i förordningen, *hög risk*, och de AI-system som faller utanför dessa kategorier. Trots detta riskbaserade angreppssätt led-

de parlamentets förslag som nämnt, via förhandlingar i trilogen, till en allmän reglering för GPAI, även om GPAI-regleringen i sin tur fick en parallell nivåindelning med en särreglering av de allra mest potenta GPAI-systemen som kan medföra ”systemiska risker”.

Utöver förbjudna praktiker kan man peka på högriskkategorin som den kanske mest praktiskt centrala för regleringen, i det att den tillåter användning men bara under vissa krav. Praktiskt sett blir *bedömningen* om en utvecklares eller vidareanvändares (”deployer”) AI-system träffas av högrisknivån eller inte väldigt viktig. I korthet avgörs det av om de träffas av något av användningsområdena som listas i bilaga III, samtidigt som de kan bedömas medföra betydande (”significant”) risk för skada för hälsa, säkerhet eller grundläggande rättigheter (enligt artikel 6.2a). AI-system som profilerar individer klassificerar EU-lagstiftaren också som högrisk-AI. Högrisk-AI-system kan enligt bilaga III gälla AI-system vid rekrytering, AI-system för att utvärdera fysiska personers kreditvärdighet eller för att fastställa deras kreditbetyg.

Lagstiftningen ställer en rad generella krav på högrisksystem, som i mycket handlar om att ha riskhantering på plats, och på att data som används för träning uppfyller vissa kvalitetskriterier. Intressant ur transparenshänseende är kraven på automatisk loggning av händelser för att möjliggöra övervakning av driften, på information till användare och på att systemet utformas på ett sätt som möjliggör mänsklig tillsyn.

Den centrala frågan kvarstår: hur ska en AI-reglering hantera taktproblemet att AI-utvecklingen går fort, och därigenom försöka reglera ett rörligt mål?

Lagstiftningen medför också en rad skyldigheter för leverantörer av högrisksystem. De ska bland annat säkerställa att AI-systemet genomgår ett förfarande för bedömning av överensstämmelse innan systemet tas i bruk eller släpps ut på marknaden, fullgöra registreringsskyldigheter, CE-märkning om överensstämmelse, upprätta kvalitetsstyrningssystem med mera.

Flexibilitet i ljuset av snabb innovation

Den centrala frågan kvarstår: hur ska en AI-reglering hantera taktproblemet att AI-utvecklingen går fort, och därigenom försöka reglera ett rörligt mål? Svaren på EU-lagstiftarens försök ligger i en rad mekanismer som syftar till en ”framtidssäker” flexibilitet. Det handlar både om riktlinjer och uppförandekoder, harmoniserade standarder, och att lagstiftningen i flera fall bemyndigat kommissionen med att, till exempel, göra tillägg till listan med högrisk-AI genom så kallade delegerade akter. Denna mekanism används även i relation till hur de ”systemiska” riskerna ska klassificeras för GPAI-modeller, vilket diskuterats ovan.

Delegerade akter beskrivs som ”akter med allmän räckvidd som inte är lagstiftningsakter och som kompletterar eller ändrar vissa icke väsentliga delar av lagstiftningsakten” (enligt artikel 290 i fördraget om Europeiska unionens funktionssätt). Ett liknande instrument finns i så kallade genomförandeakter – även om de delegerade akterna är mest betydande för AI-förordningen – som är ett verktyg i EU-lagstiftningens verktygslåda som syftar till att skapa enhetliga förutsättningar för genomförandet av den aktuella rättsakten, genom att ge befogenhet till kommissionen (eller i undantagsfall Europeiska unionens råd) att utfärda genomförandeakter (enligt artikel 291 i fördraget om Europeiska unionens funktionssätt).

Dessa är med andra ord ett slags temporal flexibilitet, där lagstiftningen kan ändras i vissa delar i framtiden. Det blir därmed av central betydelse exakt vilka delar som kan ändras. Genom de olika lagförslagen från kommissionen, rådet och parlamentet verkar delegerade akter ha varit ett sätt att hantera utmaningar med flytande begrepp och verklig osäkerhet kring vart AI-området är på väg. Enligt artikel 7 i AI-förordningen har kommissionen befogenhet att genom delegerade akter lägga till AI-system till listan över tekniker och tillvägagångssätt som specificerar högrisk-AI-system, förutsatt att dessa system också innebär en motsvarande risk för hälsa och säkerhet eller påverkan på grundläggande rättigheter.

Med samma tillvägagångssätt kan kommissionen ändra minimitrösklarna för vilken nivå av beräkningskraft som en GPAI-modell medför ”systemisk risk”, nämnt ovan. Kommissionen ges mandat att tillsammans med den nyinrättade AI-byrån ändra på både nivåer och hur bedömningen för ”systemisk risk” ska gå till. Det vill säga ändra på förutsättningarna i relation till teknisk utveckling som kan ge algoritmiska förbättringar och ökad hårdvarueffektivitet i framtiden. Även om det är förståeligt i ljuset av de principiella uppdrag AI-förordningen har

i relation till en AI-utveckling som kan ändras tämligen snabbt framöver, kan man samtidigt fråga sig vad det innebär för marknadsaktörer att på förhand inte kunna veta hur gränsnivåerna mot ”systemisk risk” kommer att stå sig. Det tycks flexibelt och svårförutsägbart på samma gång. De ökade krav som tillkommer om en GPAI bedöms medföra systemisk risk handlar om mer modellutvärdering, utökad riskvärdering, rapportering till den nyinrättade AI-byrån, om nödvändigt även till nationella myndigheter.

En annan del av ”framtidssäkrandet” ligger också i att så kallade harmoniserade standarder ska utvecklas, beställda av kommissionen att utvecklas inom de erkända europeiska standardiseringsorganisationerna CEN och CENELEC. När ett högrisksystem väl uppfyller de relevanta harmoniserade standarderna, är det tänkt, antas det överensstämma med AI-förordningen. Kritiker menar att denna förskjutning av forum riskerar att erodera den demokratiska grund lagstiftningen är tänkt att stå på, och medföra nackdelar för de intressen som inte har resurser för att bedriva påverkansarbete i standardiseringsförfaranden.

Adaptivt eller oförutsägbart?

Konsekvenserna av dessa flexibla element är dock inte utan utmaningar. Som nämnt i inledningen kan denna flexibilitet tyckas både aptitlig och gångbar på områden som utmärks av stor rörlighet och där de normativa svaren inte tycks färdiga – vilket delvis är fallet för generativ AI. Metoderna för transparens kring träningsdata kan utvecklas, hur marknaden kommer att sortera sig mellan producenter, och nedströmsleverantörer är i rörelse och några av de värsta konsekvenserna kan återstå att se. Det rör – enligt rapporter och namninsamlingar nämnda ovan – allt ifrån cybersäkerhet till existentiella frågor kopplade till generell artificiell intelligens, men även missbruk för att genom missinformation destabilisera och polarisera, samt redan konstaterade problem med bias och reproduktion av diskriminerande samhällsstrukturer (se här gärna min och mina kollegors artikel från 2023, ”Towards a Socio-Legal Robotics” där vi visar på dilemman med adaptiv teknik som tränas på orättvisa och stereotypa sociala strukturer). Baksidan på samma flexibla regleringsmynt handlar samtidigt om brist på rättslig förutsägbarhet och därmed ytterst om rättsosäkerhet.

Man kan argumentera för att alternativen är få för EU-lagstiftaren. Det vill säga att AI-fältet behöver regleras, med mer dokumentation och transparens och tydliggöranden om ansvarsfördelning, så att de omvittnade riskerna kan motarbetas. Samtidigt är AI-fältet i rörelse, pådrivet av några av de resursstarkaste aktörerna i världen, vilket påverkar vilka metoder som ska kallas ”AI”, och det är inte klart hur marknaderna och den vardagliga användningen organiseras och landar i när det gäller exempelvis generativ AI. Ändå framstår, allt sammantaget, den stig som EU-lagstiftaren slagit in på som smal, men förståelig, med inbyggda element av oförutsägbarhet som framförallt kommissionen kommer att få företräda. Viss sorts flexibilitet gör att AI-regleringen förskjuter makt över bestämmandet i lagstiftningsprocess till kommissionen, en annan till standardiseringsorgan. Den europeiska lagstiftaren tycks både vilja adressera generativ AI och samtidigt skjuta det på framtiden.

Ändå framstår, allt sammantaget, den stig som EU-lagstiftaren slagit in på som smal, men förståelig, med inbyggda element av oförutsägbarhet som framförallt kommissionen kommer att få företräda.

Avslutningsvis, vad är då att vänta för svenskt vidkommande? En kort men trolig tolkning är att tillsynen kommer att bli viktig. Det kommer att behövas vägledning i rättsliga frågeställningar kopplade till AI-utveckling och -tillämpning. Det leder samtidigt till ett behov av ett proaktivt förhållningssätt som inte alltid varit tillsynens signum. Lärandet kommer troligen att behöva ske i dubbla riktningar – både utvecklares och användares rättsliga och sociotekniska förståelse behöver stimuleras, såväl som att myndigheter måste finna sätt att säkerställa AI-relaterad kompetensutveckling, i syfte att kunna tillgodose denna proaktivitet. Myndighetssamverkan tycks central, eftersom både AI-tillämpningar generellt och AI-förordningen specifikt träffar så många sakområden. Detta är särskilt relevant med tanke på AI-förordningens ämnesmässiga bredd, med alltifrån frågor om diskriminering och grundläggande rättigheter till tekniskt avancerade frågor, som kan relatera till finansiella marknader, konkurrens, konsumenter, medicinska applikationer och dataskyddsfrågor, med mera.

Kort sagt, vi befinner oss i en formativ period i skärningen mellan normer och AI, både globalt, i Europa och i Sverige.

STEFAN LARSSON
Jurist och docent i teknik och social förändring vid institutionen för teknik och samhälle på Lunds universitet