AI-FÖRORDNINGEN – HUR EN ENKEL IDÉ BLEV TILL EN SMÖRGÅSTÅRTA AV NYANSER OCH UNDANTAG

Katja de Vries*

1. INLEDNING

Den 21 april 2021 lämnade EU-kommissionen ett 127 sidor långt förslag till förordning om Artificiell Intelligens (AI-förordningen).[1] Därefter följde nästan tre år av lobbying, politiska kompromisser och teknisk utveckling som utmanade rättsliga kategorier och tvingade fram nyanser. Industrin varnade för att regleringen skulle underminera EU:s innovationskraft, medan människorätts- och dataskyddsorganisationer menade att förordningen urvattnade rättighetsskyddet. Texten fördubblade sin storlek. Den 13:e mars 2024 röstade EU-parlamentet igenom en version av AI-förordningen[2] om 262 sidor. Blev det bra? Inom AI-industrin anser många att utfallet är besvärligt, men hanterbart – det kunde ha varit mycket värre.[3] Hos människorätts- och dataskyddsorganisationer finns fortfarande mycket besvikelse kring det urvattnade rättighetsskyddet, särskilt när det gäller användningen av biometriska AI-system, olika undantag för brottsbekämpande myndigheter, och en avsaknad av skydd för migranter,

* Katja de Vries är universitetslektor i offentlig rätt vid Uppsala universitet.

1 Förslag till Europaparlamentets och rådets förordning om harmoniserade regler för artificiell intelligens (Rättsakt om artificiell intelligens) och om ändring av vissa unionslagstiftningsakter, COM(2021) 206 final, Bryssel den 21 april 2021.

2 Europaparlamentets lagstiftningsresolution av den 13 mars 2024 om förslaget till Europaparlamentets och rådets förordning om harmoniserade regler för artificiell intelligens (rättsakt om artificiell intelligens) och om ändring av vissa unionslagstiftningsakter (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). I mitt bidrag kallar jag den här versionen den slutgiltiga, men det är möjligt att artikelnumren ändras i den officiella publicerade versionen.

3 E. Gkritsi, AI Act's global effects might be overstated, experts say, Euractiv 21 mars 2024.

flyktingar och asylsökande.[4] Dessutom anses alla nyanser och undantag göra det lättare för industrin att hitta kryphål i lagstiftningen.

Globalt sett är AI-förordningen ändå en milstolpe: Medan andra länder som Kina, USA och Storbritannien reglerar AI på ett mer vertikalt sätt (inom särskilda sektorer) eller genom allmänna principer, har EU skapat världens första övergripande och samtidigt detaljerade AI-lagstiftning. AI-förordningen är en ambitiös balansgång: den syftar till att främja EU som ledare inom AI-innovation och samtidigt skydda grundläggande rättigheter och EU:s värden. En CE-märkning som visar att ett AI-system har utvecklats och tillämpats i enlighet med förordningen ska fungera som en garanti och ett varumärke för tillförlitlig och människocentrerad AI. Förhoppningen är att förordningen kommer att ha en global inverkan genom ”Bryssel-effekten”[5]: att tillverkare utanför Unionen kommer att följa förordningen för att tillförsäkra sig tillgång till alla 460 miljoner konsumenter på EU:s marknad och att de inte kommer att orka följa olika regler i olika delen av världen och därför väljer att följa EU-reglerna.

Efter att AI-förordningen har genomgått en sista språklig och juridisk granskning (”corrigendum-proceduren”) och har fått Rådets formella godkännande, förväntas den offentliggöras i EUs officiella tidning i maj 2024, och träda i kraft 20 dagar senare. Förordningen blir fullt tillämplig 24 månader efter ikraftträdandet, men det finns undantag (art. 113). Till exempel, förbudet mot oacceptabla AI-tillämpningar börjar gälla efter 6 månader (november 2024), regler för AI-modeller av mer allmän karaktär (”general purpose AI”) efter 12 månader (maj 2025) och skyldigheter för högrisksystem efter 36 månader (maj 2027). Enligt art. 70(2) ska medlemsstaterna inom 12 månader (maj 2025) utse nationella behöriga myndigheter (art. 3(48)) som ska ansvara för marknadskontroll och tredjepartsbedömning och certifiering.

2. EN ENKEL IDÉ: RISKMODELLEN

Risk, enligt den slutgiltiga AI-förordningen, är “kombinationen av sannolikheten för att en skada inträffar och skadans allvarlighetsgrad” (art. 3(2)). AI-system kan innehålla olika nivåer av risker för hälsa och säkerhet eller grundläggande rättigheter. Kommissionens ursprungliga förslag från 2021 byggde på en enkel riskmodell som kan beskrivas i fem premisser som också formar ryggraden i den slutgiltiga versionen av förordningen.

Först, att de flesta AI-system är ganska oskyldiga och inte innehåller många risker. Förordningen är inte en lag som bara kommer med förbud, restriktioner

[4] L. Lazaro Cabrera, EU’s much-heralded AI Act agreed by EU Parliament – but serious human rights holes in law remain, Euractiv 14 mars 2024; E. Gkritsi, The long and winding road to implement the AI Act, Euractiv 14 mars 2024.

[5] Gkritsi (21 mars 2024), n. 3.

och krav. Tvärtom, ett viktigt syfte med förordningen är att AI-system med låg eller minimal risk, vilket är majoriteten av alla AI-system, kan utvecklas utan betungande regleringar och på så sätt uppmuntra AI-innovation inom EU. Exempel som nämns av EU-kommissionen[6] är AI-baserade videospel och skräppostfilter.

Andra premissen är att AI-system inte är oacceptabla eller riskabla i sig: det är bara när de tillämpas inom ett särskilt område och på ett särskilt sätt som de kan bli det. Detta uttrycks tydligt i definitionen[7] av ett AI-system (art. 3(1)): ”a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments”. Om ett system inte påverkar något faller det utanför förordningens tillämpningområde. En forskare som utvecklar ett AI-system ska inte behöva bry sig om AI-förordningen så länge systemet inte släpps ut i världen. Förordningen ska inte tillämpas på ”AI systems or AI models, including their output, specifically developed and put into service for the sole purpose of scientific research and development” (art. 2(6)) eller ”any research, testing or development activity regarding AI systems or models prior to their being placed on the market or put into service” (art. 2(8)).

Tredje premissen handlar om transparenskrav för vissa AI-system (art. 50). Om man pratar med en chatbot bör det avslöjas att det handlar om ett AI-system; om känslor eller egenskaper uttyds automatiskt bör man informeras om detta; och om ett AI-system används för att generera eller manipulera bild-, ljud- eller videoinnehåll som liknar autentiskt innehåll bör det avslöjas att innehållet har skapats på automatiserad väg.

Fjärde premissen är att det finns vissa oacceptabla tillämpningar som ska förbjudas (art. 5). Gruppen av oacceptabla tillämpningar var redan väldigt begränsad i Kommissionens förslag, och har blivit ännu mer begränsad i den slutgiltiga versionen. Den praktiska betydelsen av förbudet av oacceptabla tillämpningar som nämns i art. 5 kommer förmodligen inte att vara så väldigt stor. Förbudet i art. 5 har främst en ideologisk och politisk betydelse: att EU inte ska ha ett kinesiskt socialt kreditsystem, eller bli en dystopi med biometrisk fjärridentifiering av ansikten i realtid, system som förutsäger vem som ska begå brott i stil med *Minority Report*, eller använda AI-system som utnyttjar sårbarheter eller använder subliminala tekniker som orsakar fysisk eller psykisk skada. Problemen är att skiljelinjen mellan dystopiska tillämpningar och ”det-som-man-måste-kunna-göra!”, har visat sig vara väldigt omdebatterad. Till exempel, ponera ett system som använder biometrisk fjärridentifiering i realtid för att

[6] EU-kommissionen, AI Act, 2024, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

[7] Europaparlamentets lagstiftningsresolution om AI-förordningen (13 mars 2024), n. 2.

identifiera individer som går mot rött ljus, sedan visar upp deras namn på en stor skärm och automatiskt bestraffar genom att dra ner den sociala krediten på ett sätt som gör det omöjligt att boka tågbiljetter: Det skulle vara en oacceptabel tillämpning, eftersom den sociala poängsättningen leder till en ”skadlig eller ogynnsam behandling (…) som saknar koppling till de sammanhang i vilka berörda data ursprungligen genererades eller samlades in” och ”är omotiverad eller oproportionerlig i förhållande till personernas sociala beteende eller till hur allvarligt beteendet är” (art. 5(1c)). Dessutom är brottet inte tillräckligt allvarligt för att göra det acceptabelt att brottsbekämpande myndigheter använder sig av biometrisk fjärridentifiering i realtid. Daremot skulle det ha varit acceptabelt om den biometriska fjärridentifieringen i realtid användes för att bekämpa ett allvarligt brott (till exempel ett brott som nämns i bilaga II och kan bestraffas med fyra år i fängelse, se art. 5(1h)) och om AI-klassificieringen skulle ha varit proportionerlig och grundats på data med en relevant koppling.

Femte premissen är att AI-system som tillämpas på ett sätt som innehåller en hög risk (art. 6) för hälsa, säkerhet, eller grundläggande rättigheter bara ska tillåtas om de uppfyller särskilda krav. Det finns väsentliga krav (art. 9–15), till exempel att det ska finnas ett riskhanteringssystem (art. 9) och att särskilda krav uppfylls kring datakvalitet och dataförvaltning (art. 10), transparens och tillhandahållande av information till användare (art. 13), mänsklig tillsyn (art. 14) och noggrannhet, robusthet och cybersäkerhet (art. 15). Det finns också skydigheter för vissa aktörer. Till exempel, leverantörer av högrisksystem ska inrätta ett kvalitetsstyrningssystem (art. 17), teknisk dokumentation (art. 18), och automatiskt genererade loggar (art. 20), och på begäran av en nationell behörig myndighet förse den med all information och dokumentation som krävs för att visa att systemet uppfyller alla väsentliga krav (art. 23).

För att klassas som högrisksystem ska AI-systemet (enligt art. 6(1)) vara en säkerhetskomponent i en av produkterna som nämns i bilaga I (t.ex. AI-tillämpningar vid robotassisterad kirurgi) eller (enligt art. 6(2)) tillhöra en av de åtta tillämpningsområdena som nämns i bilaga III: biometri (t.ex. biometrisk fjärridentifiering), kritisk infrastruktur (t.ex. tillgång till el eller vatten), utbildning (t.ex. poängsättning av tentor), anställning (t.ex. CV-sorteringsprogram för rekryteringsförfaranden), viktiga privata och offentliga tjänster (t.ex. kreditpoäng som nekar medborgarna möjlighet att få ett lån), brottsbekämpning som kan inkräkta på människors grundläggande rättigheter (t.ex. utvärdering av bevisens tillförlitlighet), migration, asyl och gränskontroll (t.ex. automatisk prövning av viseringsansökningar), och rättskipning och demokratiska processer (t.ex. AI-lösningar för att söka domstolsavgöranden).[8]

[8] Kommissionen (2024), n. 6.

3. HÖG RISK ELLER INTE? FÖRORDNINGENS STORA FRÅGA

AI-förordningen innehåller 113 artiklar och 6 till 51 handlar om hög risk. Regleringen av högrisksystem utgör den största delen av AI-förordningen, och svaret på frågan om ett system medför hög risk eller inte har väldigt stora konsekvenser. Om ett system inte klassas som högrisksystem betyder det i de flesta fall att det medför minimal eller låg risk. Om ett system medför hög risk finns det väldigt många krav att uppfylla, men om det medför minimal eller låg risk finns det i princip inga krav förutom viss självreglering. Skillnaden är som dag och natt. Förhandlingarna kring vad som klassas som hög risk har varit utdragna och har lett till en komplex skiktad (”tiered”) regelgivning med olika undantag. Klassificeringen av högrisksystem enligt art. 6(1) är enkel: Ett system medför hög risk om det är en säkerhetskomponent i en av produkterna som nämns i bilaga I. Hög risk klassificeringen enligt art. 6(2) låter till en början också enkel: Alla AI-system som faller inom tillämpningområdena från bilaga III medför hög risk. Förhandlingarna kring bilaga III har varit hårda och tillämpningsområdena har formulerats snävt. Till exempel, tillämpningsområdet ”migration, asyl och gränskontroll” innehåller långt ifrån alla AI-system inom området, utan bara några väldigt specifika tillämpningar som lögndetektorer och automatiska prövningar av viseringsansökningar. Trots detta har det i den slutgiltiga versionen lagts till ännu ett tröskel: Enligt art. 6(2a) ska AI-system som faller inom bilaga III inte klassas som högrisksystem om systemet ”inte utgör en betydande risk för skada på hälsa, säkerhet eller grundläggande rättigheter, inbegripet genom att inte i väsentlig grad påverka resultatet av beslutsfattande”. Därefter nämns flera fall där det finns en presumtion att systemet inte medför hög risk: om systemet används för att (a) utföra en snäv processuell uppgift, (b) förbättra resultatet av en tidigare avslutad mänsklig aktivitet, (c) upptäcka mönster i beslutsfattandet, eller (d) utföra en förberedande uppgift för en bedömning. Till slut följer ett undantag från presumtionerna: Om AI-systemet bygger på profilering av naturliga personer gäller inte presumtionerna och systemet ska klassas som ett högrisksystem. För att inte göra det för lätt att komma undan riskklassificeringen ska undantag på grund av art. 6(2a) dokumenteras enligt art. 6(2b).

4. NEW LEGISLATIVE FRAMEWORK

AI-förordningen tillhör EU:s lagstiftningsram ”The New Legislative Framework”[9] (NLF), som har funnits sen 2008, och som i sin tur bygger på ”The New Frame-

[9] Europaparlamentets och Rådets Förordning EC 765/2008 av den 9 juli 2008 om krav för ackreditering och marknadskontroll i samband med saluföring av produkter och upphävande

work" (1985). NLF används för att lagstifta kring ackreditering och CE-märkning av produkter för EU:s inre marknad. När man köper ett gosedjur eller tar en hiss så betyder en CE-markering att produkten är skapad i konformitet med EU:s produktsäkerhetslagstiftning: Ens barn kan tugga på gosedjuret utan att bli förgiftat av kemikalier och hissen uppfyller alla säkerhetskrav. NLF lagstiftning är ett slags teknisk lagstiftning som översätts av standardiseringsorganisationer i mer detaljerade och tillämpbara standarder som tillverkare kan tillämpa. Exempel på standardiseringorganisationer är SIS (Svenska institutet för standarder) på nationell nivå, CEN (European Committee for Standardisation) och CENELEC (European Committee for Electrotechnical Standardisation) på EU nivå, och IEC (International Electrotechnical Commission) och ISO (International Standardization Organization) på internationell nivå. Olika nivåer brukar också påverka varandra: nationella organisationer sitter runt bordet på EU-nivå och internationell nivå. Till exempel har svenska regeringen utsett SIS till svenskt standardiseringsorgan inom CEN och ISO. Dessutom har CEN och CENELEC möjlighet att skapa harmoniserade standarder för AI-system (art. 40). Lagstiftning som ska vidareutvecklas i standarder är inte oproblematisk från ett demokratiskt perspektiv. När tekniska standarder skapas samlas intressenter, oftast från industrin, som betalar för att vara med i utformningsgruppen, och när en standard är färdig säljs den. Det betyder att tolkningen av produktsäkerhetslagstiftning händer bakom stängda dörrar och att det är mest industrin och privata sektorn som sitter runt bordet, och att den standard som skapas inte är fritt tillgänglig.

Standardisering kommer att ha en nyckelroll i det praktiska genomslaget och betydelsen av AI-förordningen.[10] Standarder kommer att förse leverantörerna med tekniska specifikationer för att säkerställa förordningens efterlevnad. Redan i maj 2023 skickade Kommissionen en begäran till CEN/CENELEC att ta fram europeiska standarder för AI-förordningens grundläggande krav (art. 9–15) på högrisksystem.[11] Även om det nu finns en slutgiltig version av förordningen, är det omöjligt att utvärdera hela lagstiftningen än. Det är tolkningen av förordningen i standarderna som kommer att vara styrande i praktiken: Särskilt CENs/CENELECs harmoniserade standarder som skapas på Kommissionens uppdrag, publiceras i den officiella tidningen och antas vara i överenstämmelse med förordningen.

av förordning (EEG) nr 339/93; Förordning (EU) 2019/1020 av den 20 juni 2019 om marknadskontroll och överensstämmelse för produkter och om ändring av direktiv 2004/42/EG och förordningarna (EG) nr 765/2008 och (EU) nr 305/2011.

10 S. Larsson, Mellan juridisk fixering och flexibilitet i den europeiska AI-förordningen, Advokaten, 90(2), 2024; Gkritsi (14 mars 2024), n. 3.

11 C(2023)3215 – Standardisation request M/593**.** Commission Implementing Decision of 22.5.2023 on a standardisation request to the European Committee for Standardisation and the European Committee for Electrotechnical Standardisation in support of Union policy on artificial intelligence.

Mot betalning av några tiotusen per år kan intressenter i Sverige anmäla sig att delta i utvecklingen av standarder i kommittén SIS/TK 421 Artificiell intelligens. Standardisering är en öppen process, men inte gratis, och det är ett tidskrävande frivilligt arbete. De som har anmält sig hittills är mest aktörer inom industri och privata sektorn: företag, några advokat- och konsultbyråer, och samarbetsorganisationer för forskning och näringsliv. Också på EU-nivån (kommittén CEN-CLC/JTC 21) och på internationell nivå (kommittén ISO/IEC och JTC 1/SC 42) är det mest industrin som sitter vid bordet, även om det finns vissa mindre initiativ för att främja civilsamhällets deltagande.[12]

När standarderna är färdiga kommer de i första hand att användas för självbedömning av överensstämmelsen med alla krav. Det är leverantörens eget ansvar att göra en bedömning av om ett högrisksystem är utformat i överensstämmelse med kraven (skäl 125). Bara i några få högrisktillämpningar, nämligen då AI-systemet är en säkerhetskomponent eller när det handlar om biometriska system, kommer bedömningen att utföras av en tredje part.

5. UTMANINGEN ATT LAPPA IHOP EN BLANDNING AV OLIKA RÄTTSOMRÅDEN

Som någon som har forskat länge inom EU-informationsrätt och rättighetsskydd, var jag förvånad när jag först läste Kommissionsförslaget till AI-förordning 2021. NLF-ramverket och terminologin var för mig att komma ut på okänt vatten. AI-förordningen är inte heller helt vanlig NLF lagstiftning. Även om NLF-ramverket utgör grunden till förordningen, handlar den också väldigt mycket om rättighetsskydd och cybersäkerhet. Även för en rättsexpert inom NLF-rättsområdet innehåller AI-förordningen förmodligen många oväntade delar.

Klassisk NLF-lagstiftning, till exempel direktivet om hissar,[13] riktar sig nästan helt mot tillverkare, och brukar inte ha något att göra med rättigheter. AI-förordningen däremot riktar sig först och främst mot ”leverantörer” (”en fysisk eller juridisk person, en offentlig myndighet, en byrå eller ett annat organ som utvecklar eller låter utveckla ett AI-system eller flerändamåls AI-modell i syfte att släppa ut det på marknaden eller ta det i bruk i eget namn eller under eget varumärke, antingen mot betalning eller kostnadsfritt’”, art. 3(3)) och ”användare” (”varje fysisk eller juridisk person, offentlig myndighet, byrå eller annat organ som under eget överinseende använder ett AI-system, utom när AI-systemet används inom ramen för en personlig icke-yrkesmässig verksamhet” art. 3(4)). Till exempel, om ett universitet köper ett AI-system för att

[12] European Trade Union (ETUC), Artificial Intelligence Standardisation Inclusiveness Newsletter, https://www.etuc.org/en/artificial-intelligence-standardisation-inclusiveness-newsletter.

[13] Europaparlamentets och Rådets Direktiv (EU) 2014/33 av den 26 februari 2014 om harmonisering av medlemsstaternas lagstiftning om hissar och säkerhetskomponenter till hissar.

rätta tentor, då är universitetet en ”användare” och utvecklingsföretaget ”leverantören”. Det största ansvaret ligger på leverantören, men användare har också en hel del skyldigheter. Givet att AI-förordningen sätter rättigheterna, inklusive dataskydd, i centrum för AI-regleringen skulle man kanske förvänta sig några likheter med, till exempel, Allmänna Dataskyddsförordningen.[14] I motsats till Dataskyddsförordningen, som skapar olika rättigheter för den registrerade, fanns rättigheter för den som påverkas av högrisksystem inte med i Kommissionsförslaget. Men i den slutgiltiga versionen har det skapats två artiklar som ger rättigheter till den påverkade: rätten till förklaring av individuellt beslutsfattande (art. 85) och rätten att inge klagomål till en marknadskontrollmyndighet (art. 86).

Kraven som ställs i klassisk NLF-lagstiftning brukar vara utformade för att kunna översättas till tekniska specifikationer. Det är lätt att föreställa sig ett möte mellan intressenter hos en standardiseringsorganisation som diskuterar tekniska specifikationer av kedjor som hissen[15] ska vara upphängd i. Riskerna i klassisk NLF-lagstiftning brukar vara väldigt konkreta risker för hälsa och säkerhet, till exempel risken för att hisskorgen störtar. Riskerna brukar också vara mätbara: I ett laboratorium kan man mäta hur starka kedjorna är. När det handlar om risker för kränkning av grundläggande rättigheter blir det ofta mindre konkret och mätbart. Hur skapar man en teknisk specifikation i en standard för att till exempel uttrycka om det finns tillräcklig mänsklig översyn, datakvalitet och cybersäkerhet för att undvika onödiga kränkningar av grundläggande rättigheter som dataskydd eller likabehandling? I Dataskyddsförordningen[16] finns det ett krav på att den uppgiftsansvarige i vissa fall ska utföra en konsekvensbedömning avseende dataskydd (Data Protection Impact Assessment, DPIA). Hur man ska göra en DPIA är ett väldigt omdebatterat ämne:[17] Hur ska man mäta en risk för att en behandling av personuppgifter kan leda till en kränkning av en registrerads dataskyddsrättigheter? I den slutgiltiga versionen av AI-förordningen ska särskilda användare (myndigheter; privata aktörer som tillhandahåller allmänna tjänster; och användare av system för bedömning av kreditvärdighet eller prissättning för liv- och sjukförsäkringar) på liknande sätt genomföra en Fundamental Rights Impact Assessment (FRIA, art. 27). I det tidigare nämnda exemplet där ett AI-system används för att bedöma tentor, skulle universitetet (användaren) först behöva utföra en FRIA. Hur man exakt

[14] Europaparlamentets och Rådets förordning (EU) 2016/679 av den 27 april 2016 om skydd för fysiska personer med avseende på behandling av personuppgifter och om det fria flödet av sådana uppgifter (allmän dataskyddsförordning).
[15] Direktiv 2014/33 (26 februari 2014), n. 13.
[16] Förordning 2016/679 (27 april 2016), n. 14.
[17] K. Demetzou, Data Protection Impact Assessment: A Tool for Accountability and the Unclarified Concept of “High Risk” in the General Data Protection Regulation, Computer Law & Security Review 35(6), 2019; R. Gellert, The risk-based approach to data protection, Oxford University Press, 2020.

skulle göra är inte helt tydligt än. Utöver de mätbarhetproblem som redan har pekats ut i samband med DPIA, är frågan i samband med FRIA att det i princip handlar om alla möjliga relevanta grundläggande rättigheter. Den nyinrättade European Artificial Intelligence Office[18] (”Europeiska AI-byrån”) kommer att ge vägledning genom skapandet av en mall (art. 27(5)), men det är ytterst tveksamt om man kan skapa ett formulär som passar alla tänkbara användningskontexter.

En annan skillnad mellan klassisk NLF-lagstiftning och AI-förordningen är att den sistnämnda bygger på en riskmodell där risken beror på tillämpningen. Man kan ha ett och samma system där en användning i en kontext klassificeras som hög risk, men i en annan kontext som låg risk. I jämförelse: En hiss är en hiss. Det finns inte olika regleringar beroende på om hissen används i en skola, en affär eller på en järnvägsstation.

6. GENERAL PURPOSE AI – BOMBEN I RISKMODELLEN?

2022 års framgångar i skapandet av stora generativa AI-modeller av mer allmän karaktär, till exempel stora språkmodeller som GPT-3 eller text-till-bild modeller som Dall-E, ledde till en enorm utmaning för förordningens grundläggande idé att risker skapas i tillämpningen. När kommissionsförslaget släpptes 2021 fanns i stort sett konsensus att AI-system var bra på att lösa specifika problem men att mer generell problemlösning inte var något man kunde förvänta sig i nära framtid. Ett år senare hade mycket ändrat sig. I november 2022 lanserades ChatGPT. Tidigare under 2022 hävdade Google-ingenjören Blake Lemoine att hans samtal med språkmodellen LaMDA hade övertygat honom att systemet hade ett medvetande.[19] Lemoine fick sparken och blev kritiserad av många, men ändå var Lemoines påstående ett tydligt tecken på att diskursen runt AI hade ändrats. Geoffrey Hinton[20], en av de största pionjärerna inom AI under 1980-talet, lämnade Google 2023 och sade att den nya generationen av stora språkmodeller visade att maskiner är på väg att bli mycket smartare än vad han hade trott vara möjligt. Flerändamåls-AI (”General Purpose AI”, GPAI) är vad förordningen kallar AI-modeller som har ”tränats med en stor mängd data” och har en ”betydande generalitet och kan på ett kompetent sätt utföra ett brett spektrum av distinkta olika uppgifter”, samt integreras ”i en mängd olika system eller tillämpningar” (art. 3(63) och 51). GPAI skapade stora utma-

[18] Kommissionens beslut av den 24 januari 2024 om inrättande av Europeiska byrån för artificiell intelligens, C(2024) 390 (C/2024/1459).

[19] B. Lemoine, Blake, Is LaMDA Sentient? — An Interview, Medium, 2022.

[20] W. Douglas Heaven, Geoffrey Hinton tells us why he’s now scared of the tech he helped build, MIT Technology Review, 2023.

ningar för EU:s lagstiftare. Om man följer förordningens grundläggande logik är det omöjligt att klassa GPAI som högrisksystem. Den klassificeringen gäller bara när någon tillämpar modellerna i ett högriskområde. Dessutom blir den som tillämpar modellen ”leverantör” enligt förordningen. Till exempel, om ett universitet skulle använda GPT-4 för att bedöma och ge kommentarer till tentor, då är det universitetet som är leverantör som bär de flesta skyldigheterna, medan OpenAI kan hålla sig helt utanför allt ansvar eftersom flerändamåls-modellen GPT-4 inte i sig själv har något särskilt syfte eller tillämpning. Men tänk om OpenAI tränat sin grundmodell på dåliga data: Är det verkligen universitetet som ska ansvara för det? Efter många turer skapades en särskilda krav för GPAI-system (art. 53) och lite tyngre krav för GPAI som medför en ”systemrisk” (art. 55). Det betyder att enligt den slutgiltiga AI-förordningen är det fortfarande universitetet som är systemets leverantör, men OpenAI har också har vissa skyldigheter som GPAI-leverantör. Alla leverantörer av GPAI måste tillhandahålla teknisk dokumentation, bruksanvisningar, följa upphovsrättsdirektivet, och publicera en sammanfattning av vilka träningsdata som har använts. Leverantörer av GPAI-modeller som utgör en systemrisk har vissa extra skyldigheter som, till exempel, att genomföra modellutvärderingar, spåra och rapportera allvarliga incidenter, och säkerställa cybersäkerhetsskydd.

7. VEM SKA TA HAND OM EFTERLEVNADEN OCH GENOMFÖRANDET AV AI-FÖRORDNINGEN?

AI-förordningen innefattar ett utförligt styrnings-, tillsyns- och kontrollsystem (kapitel VII, ”Styrning”, art. 64–70 och kapitel IX, ”Post-marknadskontroll, informationsutbyte och marknadsövervakning”, art. 72–94). På EU-nivån kommer Europeiska AI-byrån[21] (art. 3(47) och art. 64) att vara det ledande kompetenscentrumet och tillsynsorganet. Frågan är om byrån kan hitta tillräckligt med högkompetenta medarbetare. Månadslönen (ungefär 40 000 kronor utan skatt) har kritiserats för att vara orimligt låg i jämförelse med det som är vanligt inom AI-industrin.[22] På nationella nivån ska varje medlemsstat inom 12 månader efter förordningens ikraftträdande utse nationella behöriga myndigheter (art. 3(48) och art. 70: åtminstone en marknadskontrollmyndighet och en anmälande myndighet) som ska främja efterlevnaden och genomförandet av förordningen. En nationell myndighet från varje medlemsstat ska vara med i Europeiska AI-nämnden (art. 65) som är ett rådgivande organ till Kommissionen. Europeiska AI-byrån och Europeiska datatillsynsmannen deltar också i nämnden men har ingen rösträtt. En fråga av stor vikt för varje medlemsstat är

[21] Kommissionens Beslut av den 24 januari 2024 om inrättande av Europeiska byrån för artificiell intelligens, C(2024) 390 (C/2024/1459).
[22] C. Stokel-Walker, Regulators Need AI Expertise. They Can’t Afford It, Wired, 14 mars 2024.

vilken myndighet som ska utses till ledande nationell behörig myndighet och delta i Europeiska AI-nämnden. Europeiska datatillsynsmannen[23] har uttryckt att rollen bör ges till datatillsynsmyndigheterna i varje medlemsland. Rollen kommer att ge den berörda myndigheten en utökad budget och vikt, och Sveriges datatillsynsmyndighet (Integritetsskyddsmyndigheten) har uttryckt stort intresse.[24]

8. AVSLUTANDE REFLEKTIONER: EN SMÖRGÅSTÅRTA DÄR MAN INTE SKA GLÖMMA GRUNDLAGRET

Efter nästan tre år av förhandlingar har AI-förordningen blivit en text med väldigt många undantag och nyanser. Den grundläggande idén, att AI-risker skapas i tillämpningen, har utmanats av GPAI men till slut delvis räddats genom att skapa särskilda krav för GPAI. Sättet som GPAI klistrades in i texten skulle man kunna kritisera, som ett tecken på att ett konsekvent ramverk och en övergripande idé saknas. Samtidigt skulle man kanske också kunna applådera det, som en bra pragmatisk lösning i ett fält där anpassning och flexibilitet[25] gentemot tekniska utvecklingar behövs? Förordningen ger till exempel Kommissionen möjlighet (skäl 173) att genom delegerade akter ändra vilka tillämpningsområden som medför hög risk (art. 6(6) och 7), rekvisiten för ”GPAI med systemrisk” (art. 51(3)) och krav på teknisk dokumentation (art. 11(3)). Förutom de här ändringsmöjligheterna finns det också en annan flexibilitet: Lagtexten är bara ett första steg i regleringen och dess framgång eller misslyckande beror till stor del på utarbetandet av standarder och effektiviteten hos styrningssystemet. Det är en vinglig konstellation och det är lätt att gå vilse i en reglering som är komplex och beroende av många aktörer med olika intressen och bakgrunder. Flexibilitet, komplexitet och en blandning av olika aktörer är inte dåligt i sig, så länge man åtminstone kan samsas runt det som kanske är AI-förordningens mest grundläggande idé: att man ska reglera AI-system där det finns riktiga risker för hälsa och säkerhet eller grundläggande rättigheter.

[23] European Data Protection Supervisor (EDPS), Opinion 44/2023 on the Proposal for Artificial Intelligence Act in the light of legislative developments, 23 October 2023.
[24] Integritetsskyddsmyndigheten, AI-förordningen kommer med nya uppdrag till IMY, 14 mars 2024.
[25] S. Larsson (2024), n. 10.