



# LUND UNIVERSITY

## AI i offentlig sektor

### Från etiska riktlinjer till lagstiftning

Larsson, Stefan; Ledendal, Jonas

*Published in:*  
De Lege årsbok 2021

2022

*Document Version:*  
Förlagets slutgiltiga version

[Link to publication](#)

#### *Citation for published version (APA):*

Larsson, S., & Ledendal, J. (2022). AI i offentlig sektor: Från etiska riktlinjer till lagstiftning. I K. de Vries, & M. Dahlberg (Red.), *De Lege årsbok 2021: Law, AI & Digitalization* (s. 467-495). (De Lege årsbok; Vol. 2021). lustus förlag.

*Total number of authors:*  
2

*Creative Commons License:*  
Ospecificerad

#### **General rights**

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117  
221 00 Lund  
+46 46-222 00 00

Stefan Larsson & Jonas Ledendal

# AI i offentlig sektor: Från etiska riktlinjer till lagstiftning

## 1 Inledning

Användningen av artificiell intelligens (AI), det vill säga en rad primärt databeroende metoder och teknologier för bland annat prediktion och automation, tycks bidra till en omvälvande period i den offentliga förvaltningen. Detta skifte föranleder ett styrningsbehov, ofta kopplat till tillitsfrågor, vilket i en internationell kontext inte minst syns i en stor mängd etiska riktlinjer och principiella dokument som publicerats under de senaste få åren. Styrningen befinner sig därmed i en formativ period, vilket inte minst tydliggörs i och med EU-kommissionens förslag på den förordning om harmoniserade regler för AI (rättsakt om AI) som publicerades i april 2021.<sup>1</sup> Vi vill här belysa denna styrningsutveckling i sig, visa på dess mest centrala delar, samt analysera den svenska offentliga AI-användningen i dess ljus.

### 1.1 Regeringsuppdrag i ljuset av internationell AI-styrning

I juni 2021 fick Myndigheten för digital förvaltning (DIGG), Arbetsförmedlingen, Bolagsverket, och Skatteverket i uppdrag av regeringen att främja offentlig förvaltnings förmåga att använda artificiell intelligens (AI) i syfte att stärka Sveriges välfärd och konkurrenskraft.<sup>2</sup> DIGG, är

<sup>1</sup> Europeiska kommissionen, Förslag till Europaparlamentets och rådets förordning om harmoniserade regler för artificiell intelligens (rättsakt om artificiell intelligens) och om ändring av vissa unionslagstiftningsakter, 21.4.2021, COM(2021) 206 final.

<sup>2</sup> Regeringen (21 juni 2021) ”Uppdrag att främja offentlig förvaltnings förmåga att använda artificiell intelligens”. Diarienummer: I2021/01825.

det tänkt, ska samordna myndigheternas arbete, som rapporteras löpande med slutredovisning senast den 20 januari 2023. Uppdragsbeskrivningen hänvisar till den nationella inriktning och den ekonomiska nytta (om 140 miljarder årligen) som DIGG redogjorde för i en rapport från januari 2020 (se vidare avsnitt 2.3). Ett huvudsakligt behov som lyfts i sistnämnda rapport återkommer i uppdragsbeskrivningen för de fyra myndigheterna, nämligen att ta fram en AI-guide för offentlig förvaltning. Guiden är enligt uppdraget tänkt att beskriva de steg en verksamhet behöver ta för att använda sig av AI, inklusive hur AI ska struktureras och tillgängliggöras. Avsikten är att en sådan guide ska anpassas till relevanta *internationella rekommendationer och riktlinjer* för AI-området och utgå från *hållbar AI* i enlighet med den nationella inriktningen för AI. Regeringsuppdraget kan därmed ses bidra i ljuset av EU-kommissionens förslag till rättsakt om AI och en vidare principiell diskurs, ofta formulerad i termer av etisk styrning av AI-tillämpning. Uppdraget sätter därmed fingret på en bredare utveckling vi vill uppmärksamma här: hur AI och databeroende maskininlärning medför metod- och processförändringar för den offentliga förvaltningen. Här väcks därmed en rad styrningsfrågor i den ständigt pågående dialektiken mellan lagstiftningsbehov och tekniska landvinningar.<sup>3</sup>

Samtidigt har sålunda styrningen på AI-området kommit att ta sig uttryck i mjuka former av reglering, genom framtagandet av etiska riktlinjer och värdegrundsbaserade ställningstaganden hos såväl globala företag som internationella sammanslutningar och stater.<sup>4</sup> Hur ser då dessa internationella rekommendationer och riktlinjer ut, vilken typ av medvetenhet och kunskap kring risker med AI bygger dessa på, och hur har myndigheterna hittills influerats av dessa? Vi avser i detta kapitel att teckna de mest centrala idéer om styrningen av AI som framkommit i internationella riktlinjer de senaste åren, med särskilt fokus på EU. Mycket av detta idéinnehåll har också funnit sin väg in i nyssnämnda förslag till en rättsakt om AI, vilket vi därmed också analyserar delar av.

<sup>3</sup> För en diskussion om samspelet mellan teknikutveckling och lagstiftning på AI-området, se Larsson, S. (2021) "AI in the EU: Ethical Guidelines as a Governance Tool", i Bakardjeva Engelbrekt, Leijon, Michalski & Oxelheim (red.) *The European Union and the Technology Shift*. Cham, Switzerland: Palgrave Macmillan.

<sup>4</sup> Jfr Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI*. Berkman Klein Center Research Publication (2020-1).

## 1.2 Syfte och disposition

Syftet med detta kapitel är att i) ge en kortfattad överblick över hur AI används i svensk offentlig förvaltning, ii) beskriva vilka utmaningar som forskningen pekar på och iii) sammanfatta hur den vidare principiella styrningsdiskussionen ser ut på europeisk nivå.<sup>5</sup> Det betyder att vi analyserar aspekter av EU-kommissionens förslag till rättsakt om AI jämfört med aspekter som lyfts fram längs vägen fram till förslaget – vilka kommit till uttryck i en rad centrala publikationer på AI-området – och ett axplock av svenska myndigheters rapportering av hur AI påverkar offentlig förvaltning. Vi diskuterar här transparens som en av de mest centrala myndighetsfrågorna kopplade till AI och automatiserat beslutsfattande i offentlig förvaltning, inte minst indikerat av kommissionens lagförslag.<sup>6</sup>

Artikelnas upplägg är som följer. Först, vi ser ett behov av att börja med en diskussion kring AI-definitionen. Detta eftersom begreppsdefinitionen är helt central för både reglering och myndigheternas teknikutveckling samtidigt som ”AI” är ett begrepp som är påfallande svårt att definiera med exakthet, och dessutom rörligt i sin betydelse över tid.

Därefter, i avsnitt 2, går vi igenom ett axplock av myndighetsrapporter som är mest relevanta för kapitlets syfte. En del av de etiska och rättsliga frågor som pekas ut av både AI-expertgruppen och EU-kommissionen fångas upp av till exempel rapporter från DIGG. En sådan är *Främja den offentliga förvaltningens förmåga att använda AI*, som i början av 2020 bland annat pekade ut behovet av en ändamålsenlig rättsutveckling.<sup>7</sup> En annan sådan är en rapport som tagits fram i samverkan mellan DIGG och Lantmäteriet under 2020 som utforskar hur det offentligas processer och rutiner kan effektiviseras genom automatisering, med bibehållen eller ökad kvalitet, samtidigt som transparensen bakom beslutsfattande bibehålls eller ökar.<sup>8</sup> I rapporten presenteras bland annat ett koncept på

<sup>5</sup> Arbetet med denna artikel har i huvudsak finansierats av Vinnova i projektet Ramverk för Hållbar AI (diarienummer 2019-00848) som har koordinerats av AI Sustainability Center. Vi vill också tacka Jockum Hildén för värdefulla kommentarer på tidigare utkast.

<sup>6</sup> För en inomrättslig analys av hur AI och automatiserat beslutsfattande förhåller sig till framförallt förvaltningsrätt och dataskydd, se Ledendal, J & Larsson, S. (kommande) *Artificiell intelligens – rättsliga ramar för automatiserat beslutsfattande i offentlig förvaltning*.

<sup>7</sup> *Främja den offentliga förvaltningens förmåga att använda AI*. Delrapport i regeringsuppdraget I2019/01416/DF || I2019/01020/DF (delvis).

<sup>8</sup> DIGG och Lantmäteriet (2020) *Testa ny teknik för automatisering inom offentlig förvaltning*. I2019/03237/DF.

en förtroendemodell för automatisering i offentlig förvaltning, med fokus på transparens. Detta avseende hur ett system för automatiserade åtgärder är uppbyggt genom tydlighet i form av en deklaration av det automatiserade systemets beståndsdelar.

De etiska och rättsliga frågor som diskuteras på AI-området har ofta sin bas i en forskningsbaserad medvetenhet som vi redovisar i korthet i avsnitt 3, med särskilt fokus på transparensfrågorna. Transparens, tillsammans med frågor om ansvar och rättvisa, är också något som utgör centrala teman för unionens policyutveckling på AI-området, vilket vi redogör för i avsnitt 4. På AI-området pågår som nämnt en betydande utveckling kring styrning (*governance*) av tillämpad AI, i termer av etiska riktlinjer och principiella dokument från både statliga aktörer, civilsamhället, EU-kommissionen och företag.<sup>9</sup> En del av denna utveckling är relevant för de rörelser på området som vi undersöker i denna artikel, och tas därför upp i urval. Mest centralt för kontexten i detta kapitel är EU-kommissionens arbete med AI-frågor så som de tagit sig uttryck genom inrättandet av en AI-expertgrupp (HLEG) som under 2019 och 2020 publicerade åtminstone fyra betydande rapporter, där *Etiska riktlinjer för tillförlitlig AI*<sup>10</sup> haft särskilt märkbar påverkan på nationella strategier i Europa.<sup>11</sup> Dessa publikationer tar i mycket sitt avstamp i kommissionens s.k. AI-strategi från 2018,<sup>12</sup> och har gjort tydliga avtryck i den vitbok som kommissionen publicerade i februari 2020.<sup>13</sup> Vitboken angav indikationer i förhållande till den riskbaserade ansats som återkom i modifierad version i förslaget på rättsakt om AI, vilket vi återkommer till nedan.

I det avslutande avsnitt 5 summerar vi överblicken över AI i offentlig förvaltning i relation till de internationella rekommendationer och riktlinjer vi analyserat, med fokus på de frågor som är mest relevanta för den

<sup>9</sup> Se exv. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399, och Larsson, S. (2020) On the Governance of Artificial Intelligence through Ethics Guidelines, *Asian Journal of Law and Society* 7(3): 437–451.

<sup>10</sup> AI HLEG (2019a) *Etiska riktlinjer för tillförlitlig AI*.

<sup>11</sup> Larsson, S., Ingram Bogusz, C., & Andersson Schwarz, J. (red.) (2020) *Human-Centred AI in the EU. Trustworthiness as a strategic priority in the European Member States*. Brussels: European Liberal Forum.

<sup>12</sup> Europeiska kommissionen, Meddelande från kommissionen till Europaparlamentet, Europeiska rådet, rådet, Ekonomiska och sociala kommittén och Regionkommittén om artificiell intelligens för Europa, 25.4.2018, COM(2018) 237 final.

<sup>13</sup> Europeiska kommissionen, *Vitbok – Om artificiell intelligens – en EU-strategi för spetskompetens och förtroende*, 19.2.2020, COM(2020) 65 final.

nära förestående utvecklingen. DIGG har också nyligen pekat på att det i offentlig förvaltning finns en stor osäkerhet när det gäller digitalisering av ärendeprocesser och automatiserat beslutsfattande. Myndigheten konstaterar att det finns ”olika uppfattningar om de rättsliga förutsättningarna och det saknas tydlighet kring exempelvis frågor gällande offentlighetsprincipen, rättssäkerhet och vad som anses vara god offentlighetsstruktur vid automatiserade förfaranden”.<sup>14</sup> Detta är exempel på utmaningar som kräver mer analys och studier framöver, i takt med att automatiserat beslutsfattande och AI-understödda metoder blir vanligare inom offentlig sektor.<sup>15</sup>

### 1.3 Kort om AI-definitionens undflyende karaktär

Trots den uppmärksamhet som AI och maskininlärning fått i både media och europeiskt policyarbete råder det inte någon strikt konsensus kring hur AI bäst bör definieras. Definitionen är samtidigt central för de rapporter som försöker mäta om eller hur AI tillämpas eller kalkylera ett monetärt värde på möjlig nytta, exempelvis genom effektivisering av offentlig förvaltning. Hur man definierar avgör helt enkelt vilka värden man summerar, och därmed vilken nytta man ser. Definitionen är vidare också naturligtvis helt central för regleringen av AI. Hur man definierar har direkt bäring på vilka aktiviteter och processer som beläggs med förbud eller kravställningar.

En rad definitioner har lanserats inom såväl forskning som i myndighetsrapporter, och en stor utmaning ligger i att det rör sig om ett dynamiskt och föränderligt fält både gällande vad man konceptuellt betecknar som ”AI”, men också i termer av att teknik- och metodutvecklingen snabbt ökar i kapacitet, kvalitet och precision i sig.<sup>16</sup>

EU-kommissionen angav i ett strategiskt dokument från april 2018 en definition som pekade på hur AI ”avser system som uppvisar intelligent beteende genom att analysera sin miljö och vidta åtgärder – med viss grad

<sup>14</sup> DIGG (2021) Rättsligt stöd till offentlig förvaltning avseende digitalisering Delrapport: Beskrivning av behovet. I2021/00288 Dnr: 2021–164, s. 8.

<sup>15</sup> Jfr Ledendal och Larsson (kommande).

<sup>16</sup> Jfr diskussion i Larsson, S. (2021) ”AI in the EU: Ethical Guidelines as a Governance Tool”, i Bakardjieva Engelbrekt, Leijon, Michalski & Oxelheim (red.) *The European Union and the Technology Shift*. Cham: Palgrave Macmillan.

av självständighet – för att uppnå särskilda mål.”<sup>17</sup> Denna definition förekommer även som utgångspunkt i DIGG:s rapport från januari 2020, och SCB:s överblick från november 2020, som vi återkommer till nedan.

Komplexiteten i begreppsapparaten ledde dock den 52-hövdade AI-expertgrupp som tillsattes av kommissionen 2019 till att föra fram en utvecklad och tämligen mångfacetterad definition i ett specifikt definitionsdokument.<sup>18</sup>

Artificiella intelligenssystem (AI-system) är programvarusystem (och eventuellt även hårdvarusystem) 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 denna data 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 den omgivande miljön har påverkats av deras föregående åtgärder.

Det finns således olika aspekter att ta fasta på i en definition, där det som pekas ut av AI-expertgruppen är a) *AI-system*, som är b) mänskligt konstruerade, c) målstyrda, d) databeroende, och därmed även anpassningsbara och reaktiva med e) någon grad av agens.

Det kan härmed konstateras att det även finns en särskild utmaning i förflyttningen mellan AI som en kreativ forskningsdisciplin till ett regleringsbegrepp.<sup>19</sup> Det vill säga, att även om en mer rörlig och flytande definition haft sina syften för utvecklingen av metoder som maskininlärning, neurala nätverk och datorseende, så finns det regleringsmässiga utmaningar med ett så pass undflyende begrepp när det används för att styra

<sup>17</sup> Med tillägget ”AI-baserade system kan vara helt programvarubaserade och fungera i den virtuella världen (t.ex. röstassistenter, bildanalysprogram, sökmotorer, tal- och ansiktsgenkänningssystem), eller inbäddas i hårdvaruenheter (t.ex. avancerade robotar, självkörande bilar, drönare eller applikationer för sakernas internet)”. Europeiska kommissionen, 25.4.2018, COM(2018) 237 final.

<sup>18</sup> AI HLEG (2019) En definition av AI: Viktigaste förmågor och vetenskapliga discipliner. <https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>.

<sup>19</sup> Larsson, S. (2021) ”AI in the EU: Ethical Guidelines as a Governance Tool”, i Bakardjieva Engelbrekt, Leijon, Michalski & Oxelheim (red.) *The European Union and the Technology Shift*. Cham, Switzerland: Palgrave Macmillan.

och reglera användningen hos företag och myndigheter. Det påverkar som nämnt också utvärderingar och bedömningar av metodernas potential, vilket vi ska se nedan, för exempelvis användningsområden inom offentlig förvaltning. Det kan därmed också konstateras att AI-expertgruppen och dess verksamhet i huvudsak befinner sig på en generell nivå, medan många av de konkreta utmaningarna behöver kontextualiseras till specifika sektorer, till exempel till svensk offentlig förvaltning, eller undersektorer inom förvaltningen – det är exempelvis stor skillnad på triagering vid akutsjukvård, fördelning av försörjningsstöd i en kommun eller användning av ansiktsgenkänning i en gymnasieskola – för att man ska kunna teckna dess utmaningar tydligare.

Även den föreslagna AI-rättsakten är naturligtvis helt avhängig en reglerad definition, vilket vi återkommer till nedan. Dock, innan vi ser till den internationella utvecklingen kring medvetenhet och styrning av AI, så följer en belysning av AI-användning inom svensk offentlig förvaltning.

## 2 Hur används artificiell intelligens i offentlig förvaltning?

I detta avsnitt ges en kortfattad överblick över hur AI har hanterats strategiskt i Sverige, och visar på ett urval av rapporter som indikerar på hur AI används i offentlig förvaltning, samt några möjliga utvecklingslinjer.

### 2.1 En nationell inriktning och en AI-agenda

Den svenska *Nationell inriktning för artificiell intelligens* som publicerades av regeringen 2018 får sägas vara det som mest liknar en officiell strategi, även om det inte är det.<sup>20</sup> Vissa bedömare kallar det en nationell strategi,<sup>21</sup> intressant nog, som därmed ses som tämligen tunn jämfört med övriga nordiska strategier.<sup>22</sup>

<sup>20</sup> Vilket bland annat påpekas i Ek, I. (2021) *AI-politik för konkurrenskraft*. Rapport AU 2022:02:01. Myndigheten för tillväxtpolitiska utvärderingar och analyser (Tillväxtanalys).

<sup>21</sup> Exempelvis EU-kommissionens granskning, se Van Roy, V. (2020) *AI Watch – National strategies on Artificial Intelligence: A European perspective in 2019*, EUR 30102 EN, Publications Office of the European Union, Luxembourg, doi:10.2760/602843, JRC119974.

<sup>22</sup> Robinson, S.C. (2020) "AI policy in the Nordics. Pledging openness, transparency and trust, while expressing readiness to apply AI in society", i Larsson, S., Ingram Bogusz,



Ett antal intressenter har dock under ledning av RISE tagit fram ”AI-agendan för Sverige”, med ett 25-punktsprogram, som publicerades den 23 februari 2021 med näringsminister Ibrahim Baylan närvarande. Inte heller detta är dock en officiell strategi. Arbetet har bedrivits i sex olika spår varav offentlig verksamhet och myndigheter är ett. Agendan efterlyser exempelvis AI-rådgivare och förändringsledare för offentlig sektor, på kommunal och regional nivå, samt att AI-satsningar inom offentlig sektor behöver koordineras och drivas mot en vision.

Det finns i AI-agendan en vision om ett ”mer automatiserat och data-drivet samhälle”, ett tydligt fokus på att bryggan och samverkan mellan näringsliv och offentlig sektor behöver utvecklas, stärkas och förenklas. När det gäller forskning och utbildning kan här särskilt lyftas fram ett uttryckt behov av mer satsningar på inter- och multidisciplinär forskning med AI i fokus som kan driva både disciplinär utveckling och AI-forskningen framåt överlag. Inom ramen för detta bör, enligt agendan,<sup>23</sup> bland annat

- universiteten skapa tvärvetenskapliga utbildnings- och forskarskoleprogram;
- forskningsfinansiärer prioritera tvärvetenskapliga utlysningar; samt
- AI definieras som ett eget forskningsområde hos forskningsfinansiärer

En tämligen omfattande del av agendan redogör för rättsliga utmaningar med osäkerhet, exempelvis kring just skydd av personuppgifter. Vidare efterlyser agendan att det juridiska ansvaret i ett alltmer automatiserat beslutsfattande behöver klargöras, vilket inte minst är en viktig fråga för offentlig sektors användning av AI.<sup>24</sup>

## 2.2 Nordiska ministerrådet om nordiska kommuner

När det Nordiska Ministerrådet under 2019 intervjuade ett urval av nordiska kommuner för att bedöma statusen och vilka frågor som ansågs viktiga, konstaterade man bland annat att användningen mest befann sig på ett teststadium även om det fanns en del tillämpningar i bruk.<sup>25</sup>

C., & Andersson Schwarz, J. (red.) (2020) *Human-Centred AI in the EU. Trustworthiness as a strategic priority in the European Member States*. Brussels: European Liberal Forum.

<sup>23</sup> RISE (2020, s. 4).

<sup>24</sup> RISE (2021, s. 5).

<sup>25</sup> Nordiska Ministerrådet (2019) *Nordiske kommuners arbeid med kunstig intelligens*.

Rapporten visar på två aspekter som var viktiga för den kommunala AI-utvecklingen, där den ena har med relationen mellan offentligt och privat att göra, och den andra med hur kommunens data ofta är organiserat i interna silos. Rapporten pekar särskilt ut betydelsen av tillit i termer av att det finns en risk för att den påverkas på ett negativt sätt om kommunerna inte är tillräckligt transparenta i sin utveckling.<sup>26</sup> Nordiska ministerrådet rekommenderar att kommunerna regelbundet utbyter erfarenheter, genomför forsknings- och utvecklingsprojekt samt utvecklar nordiska etiska riktlinjer. Sistnämnda betonade just vikten av tillit, och poängterade nordiska likheter i värderingar, förvaltningskultur och medborgarnas förtroende för det offentliga.<sup>27</sup>

### 2.3 DIGG om att främja AI i offentlig förvaltning

DIGG beskriver i en rapport från januari 2020 möjliga värden på upp mot 140 miljarder kronor årligen att omfördela inom offentlig sektor, om AI:s förtjänster kan nyttjas fullt ut. Uppskattningen inkluderar inte kostnader för ett införande – och själva övergången kan förväntas innehålla en rad utmaningar i sig som innebär kostnader – och att göra den här typen av uppskattningar bygger alltid på en rad antaganden som kan vara mer eller mindre robusta, men huvudpoängen kring möjlig nytta är tydlig. En del av svårigheten ligger, återigen, i hur man definierar AI och därmed särskiljer det från exempelvis annan automation i beslutsfattande processen. Definitionen av AI, som vi såg ovan är svår att konkretisera, utgår i DIGG:s rapport från motsvarande i EU-kommissionens Samordnade plan från 2018.<sup>28</sup>

DIGG pekar bland annat på hur AI skulle kunna nyttjas för att möta samhällsutmaningar som:

- en ökad och mera jämlik tillgång till en högkvalitativ hälso- och sjukvård trots vikande ekonomiska förutsättningar och en åldrande befolkning

<sup>26</sup> Nordiska Ministerrådet (2019, s. 36).

<sup>27</sup> Vilket undersöks av Robinson, S. C. (2020). Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). *Technology in Society*, 63, 101421.

<sup>28</sup> Europeiska kommissionen, Meddelande från kommissionen till Europaparlamentet, Europeiska rådet, rådet, Ekonomiska och sociala kommittén och Regionkommittén om samordnad plan om artificiell intelligens, 7.12.2018, COM(2018) 795 final.

- en ökad tillgång till mer jämlik och högkvalitativ utbildning som utgår från den studerandes förutsättningar, trots vikande ekonomiska förutsättningar och lärarbrist
- en offentlig förvaltning som i än större utsträckning än idag försvarar principen om lika behandling och löpande synliggör och åtgärdar upptäckt diskriminering, i syfte att värna tilliten för det offentliga

Rapporten lyfter fram dels rättsliga utmaningar i termer av rättsosäkerhet, personuppgiftsbehandling, tillgång till data från andra, säkerhets- och molnfrågor, samt transparens- och insynsfrågor (bilaga F), dels etiska utmaningar i form av transparens, partiskhet och ansvar (bilaga E). Baserat på sin utredning föreslår DIGG bland annat att ett kompetenscenter med expertis inom AI behöver utvecklas, och att rättsliga förutsättningar skapas för att underlätta försöksverksamhet. Det sistnämnda är en fråga som lyfts av flertalet utredningar,<sup>29</sup> vilket bidragit till att regeringen i januari 2021 uppdrog åt DIGG att tillhandahålla rättsligt stöd till offentlig förvaltning avseende *digitalisering*, vilket ju är en än vidare terminologi. DIGG:s rapport från 2020 föreslår också framtagandet av en ”AI-guide”, vilket också kom att ingå i det förnyade uppdraget som nämns i detta kapitelns inledning.

## 2.4 SCB om AI i Sverige

Statistikmyndigheten SCB fick under 2019 i uppdrag av regeringen<sup>30</sup> att genomföra en kartläggning av användningen av artificiell intelligens inom det svenska näringslivet och offentlig sektor samt inom universitet och högskolor. Man utgick, precis som DIGG ovan från den definition som EU-kommissionen angav i sin strategi från 2018. I den rapport som publicerades i november 2020 konstateras att 10,2 % av verksamheterna inom offentlig sektor uppger att de använt AI i någon form i sin verksamhet under 2019.<sup>31</sup> Utvärderingen konstaterar att anställdas kompetens, utbildning eller erfarenhet utgör det största hindret för användning av AI, och ekar därmed ett av de behov som EU-kommissionens AI-expert-

<sup>29</sup> Jfr Digitaliseringsrättsutredningens slutbetänkande, Juridik som stöd för förvaltningens digitalisering (SOU 2018:25); Kommunutredningens slutbetänkande, Starkare kommuner – med kapacitet att klara välfärdsuppdraget (SOU 2020:8); och Öppna datautredningens delbetänkande Innovation genom information (SOU 2020:55).

<sup>30</sup> Dnr 2019/01964/D.

<sup>31</sup> SCB, Statistiska centralbyrån (2020) *Artificiell intelligens i Sverige*.

grupp propagerar för i termer av ”datafärdigheter” (”data literacy”), eller algoritm-kunnskap.

## 2.5 Riksrevisionen om automatiserat beslutsfattande hos tre myndigheter

När Riksrevisionen under 2020 granskade beslut om föräldrapenning från Försäkringskassan, beslut om årlig inkomstbeskattning av privatpersoner från Skatteverket och beslut om körkortstillstånd från Transportstyrelsen användes förvisso en smalare definition på automation än den AI-definition som DIGG, SCB såväl som EU-kommissionens AI-expertgrupp använder sig av.<sup>32</sup> Riksrevisionen poängterar rentav att inga av de automatiserade besluten för granskning fattas med hjälp av AI.<sup>33</sup> Det konstateras dock att statliga myndigheters *automatiserade* beslutsfattande har lett till ”ökad effektivitet och att grundläggande rättssäkerhetsaspekter har till viss del har förbättrats”.<sup>34</sup> I likhet med några av inspelen ovan konstateras dock problem med att det saknas en tydlig och läsbar dokumentation av den automatiserade handläggningen, det vill säga ett slags transparensproblem. Myndigheterna följer heller inte i tillräckligt hög grad upp huruvida automatiserade beslut blivit korrekta, vilket kan beskrivas som ett behov av bättre spårbarhet, vilket vi återkommer till nedan.

Med denna mer strikta definition av automation konstaterar Riksrevisionen att automatiserat beslutsfattande till viss del har förbättrat grundläggande rättssäkerhetsaspekter, i termer av att även den manuella hanteringen blivit mer enhetlig, vilket bidrar till en ökad likabehandling. Riksrevisionen anser dock att arbetet med automatiserade beslutsprocesser kan struktureras bättre och ge bättre förutsättningar för effektiva, rättssäkra och korrekta automatiserade beslut.

<sup>32</sup> Riksrevisionen (2020:22) *Automatiserat beslutsfattande i statsförvaltningen – effektivt, men kontroll och uppföljning brister*. Stockholm.

<sup>33</sup> Riksrevisionen (2020:22), s. 4.

<sup>34</sup> Ibid.

## 2.6 DIGG och Lantmäteriet om förtroende och transparens

Lantmäteriet och DIGG utredde i samverkan under 2020 aspekter av transparens och tillit i relation till AI i offentlig sektor. Uppdraget var att utforska hur det offentliga processer och rutiner kan effektiviseras genom automatisering, med bibehållen eller ökad kvalitet, samtidigt som transparensen bakom beslutsfattande bibehålls eller ökar.<sup>35</sup> Utredningen tog fram ett koncept på en s.k. förtroendemodell som samspelar väl med kommissionens och AI-expertgruppens betoning på transparens och spårbarhet för AI-system. Förtroendemodellen syftar till transparens avseende hur ett system för automatiserade åtgärder är uppbyggt genom att sträva efter

- tydlighet i form av en deklARATION av det automatiserade systemets beståndsdelar och dess förmåga att utföra uppgifter, eller kompetens, på ett riktigt, rättssäkert och effektivt sätt och
- en strukturerad, säker och öppen logg över systemets och komponenternas identiteter och versioner, innehåll med koppling till specifika ärenden eller åtgärder som utförts

Detta skulle kunna vara en möjlig utvecklingsväg för AI och automatiserat beslutsfattande i offentlig förvaltning där utvecklingsprocess och användningsområden tydliggörs, vilket är en kritisk punkt för den offentliga förvaltningens AI-användning som vi ovan har konstaterat i relation till både smarta städer och andra delar av offentlig förvaltning. Lantmäteriet och DIGG ser att förtroendemodellen för automatiserat beslutsfattande med AI ”kan vara en viktig del för att bibehålla förtroende och tillit till den svenska offentliga förvaltningen men också till den europeiska och globala digitaliserade offentliga förvaltningen”<sup>36</sup> och rekommenderar att det utvecklas och utökas.

Denna förtroendemodell återkommer så också i ovan nämnda regeringsuppdrag som ett flertal myndigheter fick från regeringen i juni 2021.<sup>37</sup> Enligt uppdraget ska den utvecklas ”för att möjliggöra att den

<sup>35</sup> DIGG och Lantmäteriet (2020) *Testa ny teknik för automatisering inom offentlig förvaltning*. I2019/03237/DE.

<sup>36</sup> DIGG och Lantmäteriet (2020), s. 3.

<sup>37</sup> Regeringen (21 juni 2021) ”Uppdrag att främja offentlig förvaltnings förmåga att använda artificiell intelligens”. Diarienummer: I2021/01825.

kan användas som ett frivilligt ramverk för offentlig förvaltning”.<sup>38</sup> Rapporten nämner även möjligheten med ett nationellt öppet AI-register, som underlag för hur AI och automatiskt beslutsfattande används inom offentlig förvaltning, men konstaterar att det behöver utredas mer.

### 3 Forskningsbaserad medvetenhet om AI-baserade risker

Många av de rättsligt relevanta frågorna för styrningen av AI och automatiserat beslutsfattande har vuxit fram i takt med teknologiska möjligheter och forskningsbaserade insikter om både möjligheter och risker. För att förstå vilka vägval som gjorts av till exempel EU-kommissionen och dess AI-expertgrupp behöver man se vilken underliggande medvetenhet och kunskap om relationen mellan AI och samhälle som vuxit fram genom denna typ av forskning. Det är också denna underliggande kunskap som även kan bistå med att vägleda svensk offentlig sektor, vid alla de detaljerade val som behöver göras när AI-metoder implementeras och utvärderas.<sup>39</sup> Det föranleder att en viss överblick ges även i detta avseende.

#### 3.1 Medvetandegörande studier, och principiella ramverk

Det finns en rad exempel från framför allt amerikansk offentlig förvaltning som har problematiserats och studerats på ett sätt som lett till ökad medvetenhet kring risker för diskriminering. Ett sådant är det s.k. COMPAS-systemet, ett system som används i vissa domstolar för bedömning av återfallsrisk hos dömda gärningspersoner. Systemet bedömdes dock i en granskning på felaktiga grunder vara mer benäget att se högre risk hos gärningspersoner av afro-amerikansk bakgrund.<sup>40</sup>

Ansiktsgenkänning är ett annat fält som varit särskilt omdebatterat, både ur integritetsskyddsperspektiv såväl som utifrån diskriminerings- och s.k. bias-perspektiv, vilket inte minst är tydligt gällande kommissionens

<sup>38</sup> Regeringen (21 juni 2021), s. 3.

<sup>39</sup> Jfr DIGG (2020), kap. 8 och bilagor E och F.

<sup>40</sup> ProPublica (2016). ”Machine bias”, av Angwin, J., Larson, J., Mattu, S., & Kirchner, L.; jfr Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin’s Press.

förslag till rättsakt om AI.<sup>41</sup> Två amerikanska forskare vid MIT analyserade i en studie publicerad 2018 tre kommersiella system för ansiktsigenkänning. De kunde konstatera att samtliga hade sämre precision för kvinnliga ansikten och personer med mörk hy.<sup>42</sup> Något som även påvisats i system som ska uppträcka fotgängare i trafiken.<sup>43</sup> En typ av brist på precision som också tycks ha lett till högre andel felaktiga arresteringar av mörkhyade i USA,<sup>44</sup> och att flera av teknikleverantörerna upphört med att erbjuda den här typen av AI-lösningar för polisära ändamål. Liknande farhågor har visat sig med diskriminering i system för platsannonser<sup>45</sup> och digital marknadsföring,<sup>46</sup> vilket också kan ställas mot studier som visar hur svårt det är att granska användandet av AI-system och algoritmer för prediktion som utvecklas av privat sektor åt den offentliga förvaltningen.<sup>47</sup>

Denna typ av forskning har lett till insikter om att flervetenskapliga miljöer behövs, och etablerandet av internationella vetenskapliga konferenser, där en av de mer profilerade är FAccT (tidigare FAT).<sup>48</sup> Utvecklingen och medvetenheten kring risker i samspelet mellan AI-teknologier och samhälle syns också i hur traditionella vetenskapliga AI-konferenser

<sup>41</sup> Den europeiska dataskyddsstyrelsen (EDPB) och datatillsynsmannen har till exempel varnat för riskerna med fjärrbiometrisk identifiering av individer i publik miljö, och har i en gemensam åsiktsförklaring anmodan ett allmänt förbud mot det, se EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Se här även Liane Colonnas bidrag i denna volym: The AI Regulation and Higher Education: Preliminary Observations and Critical Perspectives.

<sup>42</sup> Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77–91), PMLR.

<sup>43</sup> Wilson, B., Hoffman, J., & Morgenstern, J. (2019). Predictive inequity in object detection. *arXiv preprint arXiv:1902.11097*.

<sup>44</sup> The New York Times (24 juni 2020) "Wrongfully Accused by an Algorithm in what may be the first known case of its kind, a faulty facial recognition match led to a Michigan man's arrest for a crime he did not commit", av Kashmir Hill.

<sup>45</sup> Datta, A., Datta, A., Makagon, J., Mulligan, D. K., & Tschantz, M. C. (2018, January). Discrimination in online advertising: A multidisciplinary inquiry. In *Conference on Fairness, Accountability and Transparency* (pp. 20–34). PMLR.

<sup>46</sup> Latanya Sweeney. Discrimination in online ad delivery. *Commun. ACM*, 56(5):44–54, May 2013.

<sup>47</sup> Brauneis, R., & Goodman, E. P. (2018). Algorithmic transparency for the smart city. *Yale JL & Tech.*, 20, 103.

<sup>48</sup> ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT): <https://facctconference.org>.

tämligen nyligen även börjat adressera etik och samhällsfrågor.<sup>49</sup> För svenskt vidkommande har en inriktning mot humaniora och samhälle i det 10-åriga, AI-inriktade, forskningsprogrammet WASP-HS utmärkt sig.<sup>50</sup>

I linje med detta har även utvecklingen tagit fart kring principiella ställningstaganden, där en rad aktörer från både näringsliv, offentlig sektor och civilsamhället har tagit fram etiska eller rättighetsbaserade riktlinjer för AI-utveckling och -tillämpning.<sup>51</sup> När en forskargrupp vid ett schweiziskt forskningsuniversitet under 2019 analyserade ett stort antal AI-riktlinjer identifierade de inte mindre än 84 dokument som innehöll etiska principer eller riktlinjer för AI.<sup>52</sup> En stor majoritet av dessa, 88 procent, hade publicerats från 2016 till april 2019. De undersökta dokumenten uppvisade konvergens kring de fem etiska principerna för (1) transparens, (2) rättvisa, eller det svåröversatta ”fairness”, (3) icke-skadlighet, (4) ansvar och (5) integritet. De noterade dock också att det verkar finnas stora skillnader i hur dessa principer tolkas, varför de anses viktiga, vilken fråga, domän eller vilka aktörer de relaterar till och hur de ska implementeras.

## 3.2 Transparensens förtjänster och mångfacetterade betydelse

I principiella ställningstaganden kring att tillämpa AI är behovet av transparens en av de vanligast utpekade principerna. I nämnda kartläggning av riktlinjer pekade 73 av de 84 källorna på just behovet av transparens.<sup>53</sup> En forskningsgenomgång visar dock att begreppet är mångfacetterat, inte minst med tanke på hur olika intressen såsom tillsynsbehov, användarnas informationsfärdigheter (”litteracitet”), företagshemligheter och säkerhetsfrågor behöver balanseras för en teknologi som i sig själv kan ha inne-

<sup>49</sup> Som i att AAAI / ACM utvecklat ett spår om ”Artificial Intelligence, Ethics and Society”.

<sup>50</sup> The Wallenberg AI, Autonomous Systems and Software Program – Humanities and Society: <https://wasp-hs.org>.

<sup>51</sup> Jfr Fjeld et al. (2020).

<sup>52</sup> Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nat Mach Intell* 1, 389–399 (2019). <https://doi.org/10.1038/s42256-019-0088-2>.

<sup>53</sup> Jobin, Ienca & Vayena (2019), p. 391.



boende förklarbarhetsproblem.<sup>54</sup> Även EU-kommissionen har i en rapport om säkerhets- och ansvarighetsfrågor, som publicerades samtidigt som vitboken i februari 2020, pekat ut bristen på transparens, och vad de kallar en ”black box effect” som en utmaning för rättslig efterlevnad och ansvarsutkrävande.<sup>55</sup>

Vad det är som behöver vara mer transparent behöver därmed också diskuteras i förhållande till olika aspekter av ett AI-system. Till exempel listar en rapport om ”algoritmisk transparens”, från europaparlamentets forskningservice, sju åtskilda punkter om vad som behöver åtgärdas. En av dessa riktar sig specifikt till algoritmerna i sig, medan de andra sex behandlar frågor om data, mål, resultat, efterlevnad, inflytande och användning.<sup>56</sup> EU-kommissionens AI-expertgrupp, som vi återkommer till nedan, pekar ut förklarbarhet som en av fyra grundläggande etiska principer att respektera, och transparens som ett av sju krav som AI-system bör uppfylla. Utmaningar med s.k. ”black box”-algoritmer, menar expertgruppen, behöver särskilt uppmärksammas.<sup>57</sup> De lyfter fram behovet av spårbarhet, förklarbarhet och kommunikation. Sistnämnda inkluderar både individer som interagerar med AI-system eller nås av dess utfall, men även de som yrkesmässigt interagerar med AI-system. Detta, menar vi, är särskilt relevant för området för det här kapitlet, det vill säga där handläggare i den offentliga förvaltningen kommer att interagera med exempelvis beslutsstödsystem och samtidigt ansvara för myndighetsbeslut som de fattar under påverkan av dessa stödsystem.

Aspekter av transparens av intresse för offentlig förvaltning är också hur sådan öppenhet ska säkerställas i relation till privata utförare eller utvecklare av system som de senare kan vilja hävda ska vara ”proprietära”, det vill säga deras immateriella egendom eller utgöra företagshemligheter.

<sup>54</sup> Larsson, S. (2019) The Socio-Legal Relevance of Artificial Intelligence, ”Law in an Algorithmic World”, Special Issue of *Droit et Société*. 103(3): 573–593; Larsson, S. & Heintz, F. (2020) Transparency in artificial intelligence, *Internet Policy Review* 9(2): 1–16.

<sup>55</sup> Europeiska kommissionen, Rapport från kommissionen till Europaparlamentet, rådet, Ekonomiska och sociala kommittén om Konsekvenser för säkerhet och ansvar när det gäller artificiell intelligens, sakernas internet och robotteknik, 19.2.2020, COM(2020) 64 final, s. 6–7.

<sup>56</sup> Koene, A., Clifton, C., Hatada, Y., Webb, H., & Richardson, R. (2019). A governance framework for algorithmic accountability and transparency (Study No. PE 624.262) Panel for the Future of Science and Technology, Scientific Foresight Unit (STOA), European Parliamentary Research Service.

<sup>57</sup> AI HLEG (2019a) s. 14.

ter.<sup>58</sup> Larsson och Heintz har analyserat transparensbegreppet i relation till AI utifrån sju intressen som delvis kan vara motstående.<sup>59</sup> Dessa är:

1. Förklarbarhet och utmaningar av s.k. black box-karaktär för AI-system;
2. Rättsliga aspekter av ägande eller företagshemligheter;
3. Behovet av att undvika missbruk eller ”gaming” (som en oönskad konsekvens av öppenhet);<sup>60</sup>
4. Användarnas litteracitet och kunskapsnivå;
5. Språklig eller metaforisk *översättning* av matematiska samband, exv. i användaravtal eller beslut;<sup>61</sup>
6. Marknadskomplexitet, exempelvis i så kallade dataekosystem;
7. Individuellt distribuerat utfall som en tillsynsutmaning.<sup>62</sup>

Avslutningsvis kan man konstatera att AI-transparens både kan avse förståelsen och förklarbarheten hos individuella beslut (utfall), vilket därmed kan ha olika adressater,<sup>63</sup> men också olika aspekter av hur hela

<sup>58</sup> Vilket studeras inom ramarna för s.k. smarta städer av ovan nämnda Brauneis & Goodman (2018), och i förhållande till alltmer automatiserade, datadrivna marknader av den amerikanske rättsvetaren Frank Pasquale (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

<sup>59</sup> Larsson & Heintz (2020), framförallt utvecklat i Larsson (2019).

<sup>60</sup> Vilket ur ett förvaltningsperspektiv i en rapport från Inspektionen för socialförsäkringen (ISF) förs fram av Försäkringskassan i relation till hur transparenta de bör vara med vilka variabler som ingår i hur de genomför riktade kontroller; Inspektionen för socialförsäkringen (2018) *Profilering som urvalsmetod för riktade kontroller. En granskning av träffsäkerheten, effektiviteten och rättssäkerheten i Försäkringskassans modeller för riskbaserade kontroller*. Rapport 2018:5, se exv. s. 86.

<sup>61</sup> Vilket är direkt relevant för automatiserade myndighetsbeslut. Hur de skrivs är rimligen en avgörande faktor för hur den underliggande behandlingen förstås. Vi återkommer till detta i kap. 4 nedan, inte minst i relation till dataskyddets krav på att ge ”meningsfull information om den bakomliggande logiken”, beskrivet i 4.5.

<sup>62</sup> Här avses s.k. personalisering i storskaliga automatiserade system, exv. gällande riktad reklam, och tillsynsmyndigheternas metodutmaning med att kunna granska eventuella otillbörligheter i detta, se exv. Larsson, S. (2018) *Algorithmic Governance and the Need for Consumer Empowerment in Data-driven Markets*, *Internet Policy Review* 7(2):1–12.

<sup>63</sup> Att AI-systems förklarbarhet behöver relateras till olika adressater (”audiences”) är tydligt uppmärksammat i den s.k. förklarbarhetsforskningen (xAI), jfr Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115.

AI-systemet eller beslutsstödet i sig fungerar. Sistnämnda kan i en förvaltningskontext till exempel handla om krav på ”AI-register”, med en typ av öppenhet som finns exempel på i Helsingfors och Amsterdam,<sup>64</sup> eller en typ av ovan nämnda förtroendemodellkort, som Lantmäteriet tagit fram tillsammans med DIGG under 2020.

## 4 AI i EU: etiska riktlinjer och styrning

I det här avsnittet behandlas hur EU:s policyarbete inom AI-området har utvecklats sedan 2018, genom flera betydelsefulla publikationer, se Figur 4.1 för ett urval: (1) en AI-strategi publicerad i april 2018;<sup>65</sup> (2) en anmodan publicerad i december 2018, riktad till medlemsstaterna att utveckla AI-strategier till mitten av 2019;<sup>66</sup> tillsättande av en expertgrupp (AI HLEG) som bland annat publicerat de (3) Etiska riktlinjer för tillförlitlig AI i april 2019<sup>67</sup> som fått ett relativt stort genomslag i europeiska AI-strategier;<sup>68</sup> efterföljande (4) policy- och investeringsrekommendationer för tillförlitlig artificiell intelligens,<sup>69</sup> samt den (5) vitbok som EU-kommissionen publicerade i februari 2020.<sup>70</sup> Dessa är alla av intresse för att bättre förstå det förslag till rättsakt om AI som kommissionen publicerade i april 2021 och som högst troligt även kommer att ha avsevärd betydelse för regleringen av AI i offentlig sektor. De två avslutande rapporterna från AI-expertgruppen är också intressanta och av olika karaktär: där den ena är en (6) utvecklad självvärderingsmetodik för de som implementerar eller har implementerat AI-system i sin verksamhet,<sup>71</sup>

<sup>64</sup> Analyserat av teknikfilosof Luciano Floridi, som också var medlem i AI HLEG, Floridi, L. (2020). Artificial Intelligence as a Public Service: Learning from Amsterdam and Helsinki. *Philosophy & Technology*, 33(4), 541–546.

<sup>65</sup> Europeiska kommissionen, 25.4.2018, COM(2018) 237 final.

<sup>66</sup> Europeiska kommissionen, 7.12.2018, COM(2018) 795 final.

<sup>67</sup> Publicerade samtidigt som Europeiska kommissionen, Meddelande från kommissionen till Europaparlamentet, rådet och Ekonomiska och sociala kommittén om att skapa förtroende för människocentrerad artificiell intelligens, 8.4.2019, COM(2019) 168 final.

<sup>68</sup> Larsson, S., Ingram Bogusz, C., & Andersson Schwarz, J. (red.) (2020) *Human-Centred AI in the EU. Trustworthiness as a strategic priority in the European Member States*. Brussels: European Liberal Forum.

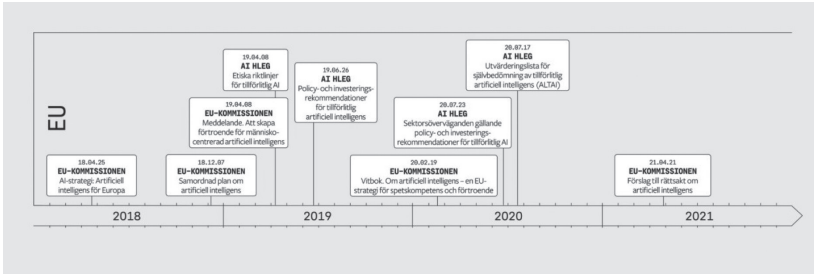
<sup>69</sup> AI HLEG (26 juni 2019) *Policy- och investeringsrekommendationer för tillförlitlig artificiell intelligens*. Bryssel.

<sup>70</sup> Europeiska kommissionen, 19.2.2020, COM(2020) 65 final.

<sup>71</sup> AI HLEG (2020) Utvärderingslista för självvärdering av tillförlitlig artificiell intelligens (ALTAI) [författarnas översättning].

och den andra handlar om (7) överväganden för specifika sektorer, vilket innefattar offentlig sektor.<sup>72</sup> Dessa är baserade på tidigare policy- och investeringsrekommendationer som har bearbetats i workshops under våren 2020 med inbjudna experter och ”stakeholders” (intressenter).

I det följande redogör vi kortfattat för 3–7 i syfte att bättre teckna dels den grund av ”internationella rekommendationer” som DIGG m.fl. myndigheter har att ta hänsyn i inom ovan nämnda uppdrag,<sup>73</sup> dels för att bättre teckna betydelsen av AI-rättsakten för svensk offentlig sektor.



Figur 4.1: Policyutveckling mot tillförlitlig AI i EU, i översättning och bearbetning från Larsson et al., 2020.

## 4.1 Etiska riktlinjer för tillförlitlig AI

HLEG’s Etiska riktlinjer för tillförlitlig AI publicerades i april 2019 som en följd av det uppdrag som kommissionen givit gruppen. Riktlinjerna består av fyra nivåer:

- i. en ram som anger att tillförlitlig AI är laglig, etisk och robust;
- ii. etiska grunder för tillförlitlig AI som finns i respekten för mänsklig autonomi, förebyggande av skada, rättvisa och ”förklarbarhet”;
- iii. sju krav för att realisera tillförlitlig AI, se Figur 4.2, samt
- iv. en utvärderingslista, direkt organiserad enligt dessa sju krav. Denna utvärderingslista testades under andra halvåret 2019 och publicerades i en uppdaterad version i juli 2020 som ”utvärderingslista för självvärdering av tillförlitlig AI (ALTAI)”, avsedd för självutvärderingsändamål.

<sup>72</sup> AI HLEG (2020) *Sektoröverväganden gällande policy- och investeringsrekommendationer för tillförlitlig artificiell intelligens* [författarnas översättning].

<sup>73</sup> Regeringen (21 juni 2021) ”Uppdrag att främja offentlig förvaltnings förmåga att använda artificiell intelligens”. Diarienummer: I2021/01825.

Nivå iii, se Figur 4.2, tycks hittills ha fått mest betydelse för förståelsen av utmaningar i europeiska nationella AI-strategier, vilket är tydligt bland annat i den polska<sup>74</sup> och den norska.<sup>75</sup> Intressant nog är vissa av dessa krav redan tydligt reglerade i rättsordningen, däribland dataskydd och kravet på ickediskriminering. Andra krav, som det om transparens, tycks helt centralt för tillförlitlighetsfrågorna, men är i sin breda betydelse inte lika tydligt reglerat för implementeringen av AI-system i alla sektorer. För den offentliga förvaltningen är dock till exempel den öppenhet som dataskyddsrätten kräver samt rätten till partsinsyn helt relevanta reglerade aspekter av transparens.<sup>76</sup>



Figur 4.2: Sju krav för realiseringen av tillförlitlig AI, från AI HLEG 2019, bearbetad i svensk översättning i Larsson, S. (2020) "AI i EU: etiska riktlinjer som styrmedel".

<sup>74</sup> Söderlund, K. (2010) "AI policy in Poland. Ethical considerations already at the core", i Larsson, Ingram Bogusz & Andersson Schwarz, red. (2020) *Human-Centred AI in the EU. Trustworthiness as a strategic priority in the European Member States*. Brussels: European Liberal Forum.

<sup>75</sup> af Malmberg, F. (2020) "AI policy in Norway. Looking to the future and harmonised with the EU", i Larsson, Ingram Bogusz, & Andersson Schwarz, red. (2020) *Human-Centred AI in the EU. Trustworthiness as a strategic priority in the European Member States*. Brussels: European Liberal Forum.

<sup>76</sup> Se Ledendal & Larsson (kommande).

#### *4.1.1 Policy- och investeringsrekommendationer*

Policy- och investeringsrekommendationerna var AI-expertgruppens andra leverans och publicerades den 26 juni 2019. Den omfattar 33 huvudpunkter (med många fler delpunkter) uppdelade i åtta grupper med rekommendationer om

- a) mänskligt bemyndigande och skyddsaspekter,
- b) transformationen av privat sektor,
- c) offentlig sektor som katalysator för hållbar tillväxt,
- d) forskningskapacitet,
- e) datahantering och infrastrukturella frågor,
- f) utbildningsfrågor,
- g) styrnings- och regleringsfrågor, och
- h) finansieringsfrågor.

Rekommendationen är detaljerad och ger vägledning för många väldigt olika sektorer och ämnen. Av störst relevans här är c och g. Expertgruppen betonar vikten av att använda de etiska riktlinjerna för tillförlitlig AI i offentlig förvaltning (12.1). Den avhandlar även rekommendationer om att individer ska kunna begära att få ”interagera” med en mänsklig handläggare om handläggningen krånglar och medför betydande påverkan på individen (9.2, se även 12.2). Detta ekar dataskyddsrättsliga frågor (se även 27.5 för explicit referens till GDPR). Expertgruppen pekar bland annat på frågor av bredare digitaliseringskaraktär, där data bör finnas digitalt (10.1), och kompetensfrågor i termer av ”datafärdigheter” för myndigheter (10.2).

De pekar också på betydelsen av offentlig upphandling (11.1–3), vilket är ett område vi också ser som särskilt centralt för frågor om AI i offentlig förvaltning, men inte fokuserar i detta kapitel. Intressant nog efterlyser de även granskningsverktyg som kan upptäcka bias och oegentligheter i myndigheternas beslutsfattande (12.3). AI-expertgruppen efterlyser bland annat en översyn av rådande reglering (28.1), utvecklandet av granskningsmekanismer för AI-system (29.4), och en riskbaserad regleringsinriktning (26.1), vilket också fokuseras i kommissionens vitbok.

## 4.2 Kommissionens vitbok

När vitboken publicerades den 19 februari 2020 åtföljdes den av en rapport om säkerhets- och ansvarseffekterna av AI, IoT och robotik (samt en europeisk datastrategi).<sup>77</sup> Som anges i vitboken är många av de frågor som de etiska riktlinjerna för tillförlitlig AI pekar på redan reglerade, till exempel inom dataskydd och antidiskriminering. Rapporten om säkerhet och ansvar diskuterar konsekvenserna av autonomi och självlärande funktioner hos AI-produkter, särskilt när det gäller riskbedömning. Detta är uppenbarligen av relevans för begreppet *människocentrerad AI*. Dessutom påpekas den brist på transparens och ”black box-effekt” som vissa AI-system kan ha på beslutsprocessen som ett verkställighets- och ansvarsproblem.

Definitionsfrågan, som vi diskuterade ovan, konstateras även här vara central för regleringsfrågorna i termer av att för ett nytt rättsligt instrument måste definitionen vara ”tillräckligt flexibel för att kunna anpassas till den tekniska utvecklingen och samtidigt vara tillräckligt exakt för att ge den rättssäkerhet som krävs”.<sup>78</sup>

Vitboken består av två huvudblock baserade på begreppet ”ekosystem”; ett om excellens och ett om tillit. Det betyder att det finns en dubbelhet i strategin: att undersöka möjligheterna å ena sidan – kopplade till efterfrågan på forskning, samarbete mellan medlemsstaterna, innovation och ökade investeringar – och riskerna eller utmaningarna å andra sidan – för att säkerställa tillförlitlighet, ansvar och säkerhet. Kommissionen konstaterar att lagstiftningsramen kan behöva utvecklas rörande exempelvis:

- effektiv tillämpning och verkställighet av befintlig EU- och nationell lagstiftning. Det vill säga att befintlig lagstiftning i många fall är ändamålsenlig men utmanas i dess implementering. Kommissionen påpekar särskilt bristen på transparens, som gör det svårt att identifiera och bevisa eventuella överträdelser;
- begränsningen av produktsäkerhetslagstiftningens tillämpningsområde som gäller produkter och inte tjänster, och därför i princip inte tjänster baserade på AI-teknik;
- föränderligheten hos AI-system, till exempel för produkter som baseras på maskininlärningsberoende programuppdateringar;
- ansvarsfördelning på olika platser i en försörjningskedja.

<sup>77</sup> Europeiska kommissionen, 19.2.2020, COM(2020) 64 final.

<sup>78</sup> Europeiska kommissionen, 19.2.2020, COM(2020) 65 final, s. 18.

- behov av utveckling av säkerhetsbegreppet, relaterat till exempelvis cybersäkerhet.

Av särskild relevans, och även föremål för debatt, är den föreslagna riskdefinitionen, eftersom den används för att identifiera behovet av framtida reglering. Kommissionen föreslår i vitboken en kumulativ ansats där den dels pekar ut högrisksektorer, till exempel hälso- och sjukvård, transport, energi och delar av den offentliga sektorn, och dels att AI-tillämpningen inom sektorn i fråga används på ett sådant sätt att betydande risker sannolikt uppstår.

Riskdefinitionen har dock fått kritik, eftersom vissa typer av risker inte nödvändigtvis uppstår i högrisksektorer, exempelvis med hänvisning till plattformsmarknader med riktad marknadsföring och sökmotorer, som därmed skulle räknas som AI-system som, trots sina risker, klassificeras som låg risk enligt kommissionens definition.<sup>79</sup> Den tyska regeringen har till exempel krävt att det riskklassificeringssystemet som föreslås i vitboken revideras, vilket i viss mån också kom att göras i det efterföljande förslaget till rättsakt om AI, från en binär ansats till en som har flera nivåer (se nedan).<sup>80</sup>

### 4.3 Sektoröverväganden och självvärdering

Som nämnt, AI-expertgruppens policy- och investeringsrekommendationer bearbetades under våren 2020 och publicerades med inriktning mot mer konkreta sektorer i slutet av sommaren 2020, varav offentlig förvaltning var en.<sup>81</sup> Expertgruppen uttrycker här behovet av att "AI-möjliggjorda e-förvaltningstjänster" bör åtföljas av adekvata arrangemang när det gäller ansvarighet och spårbarhet, för att möjliggöra efterhandsverifiering, och hänvisar här även till rätten till god förvaltning, som stadgas i artikel 41 i EU:s stadga om grundläggande rättigheter. Expertgruppen ser även några särskilda behov, exempelvis att vidta åtgärder för att främja data- och algoritm-kunnighet (notera att "datafärdigheter" – "data

<sup>79</sup> Dignum, V., Muller, C. & Theodorou, A. (2020). First Analysis of the EU Whitepaper on AI. ALLAI; se även Liane Colonnas bidrag i denna volym i relation till högre utbildning och AI.

<sup>80</sup> Die Bundesregierung (2020). Stellungnahme der Bundesregierung der Bundesrepublik Deutschland zum Weißbuch zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen.

<sup>81</sup> AI HLEG (2020) *Sektorsöverväganden*.



literacy” – har breddats) inom den offentliga förvaltningen, samt att utveckla AI- och datastrategier inom alla relevanta grenar av statlig förvaltning. I augusti 2020 publicerades också ALTAI, det vill säga expertgruppens slutgiltiga version av en utvärderingslista för tillförlitlig artificiell intelligens, som pekar på en allt viktigare utvärderingsutveckling som vi dock av utrymmesskäl inte redogör för mer här. Listan följer den sjudeklade inriktning för realisering av tillförlitlig AI som först publicerades i de etiska riktlinjerna, här kompletterade med en rad frågor.

#### 4.4 Förslag på ny AI-lagstiftning

Den 21 april 2021 publicerade kommissionen ett förslag till en förordning som syftar till att harmonisera regleringen av artificiell intelligens (rättsakt om AI). Man vill bland annat säkerställa att AI-system som släpps ut och används är säkra och är förenliga med befintlig lagstiftning om de grundläggande rättigheterna och unionens värden. Eftersom den föreslagna regleringen är i form av en förordning blir den till alla delar bindande och direkt tillämplig i varje medlemsstat och ska inte genomföras i svensk rätt, även om vissa kompletteringar kan komma att behöva göras. Den föreslagna rättsakten har naturligtvis inte uppstått i ett politiskt eller kunskapsmässigt vakuum, vilket den ovan beskrivna utvecklingen visar, men relationen till redan befintlig rätt, exempelvis på dataskyddsområdet, väcker en rad frågor. Här finns dock bara utrymme för en kort översikt, med fokus på det som är allra mest relevant för offentlig förvaltning, varför vi enbart redogör för det riskbaserade angreppssättet, hur centralt transparens är och hur man i rättsakten har hanterat AI-definitionen.

##### 4.4.1 Risknivåer som styrande mekanism

Det riskbaserade angreppssättet känns igen från vitboken, men har utökats till att inkludera flera kategorier, från förbud av AI-system med oacceptabel risk, till de med hög risk som kan vara tillåtna under vissa restriktioner och krav, de som innebär låg risk som medför transparensförpliktelser, samt de med minimal risk utan krav. Det ställs krav på både tillhandahållare och användare. I korthet presenteras:

- *Förbud* mot AI-system med oacceptabel risk, som i) använder ”subliminala metoder” för att manipulera en persons beteende på ett sätt som innebär skada; ii) utnyttjar sårbarheter hos en specifik (utsatt)

grupp för att väsentligt förändra beteendet som innebär eller sannolikt kan innebära (fysisk eller psykisk) skada; iii) profilerar eller klassificerar personers trovärdighet baserat på beteende eller personlighetsdrag om åtgärderna leder till skadlig eller ogynnsam behandling som är obefogad eller oproportionerlig i förhållande till beteendet; iv) biometriskt identifierar personer på distans i realtid på offentliga platser (med vissa möjliga undantag). I den här kategorin finns möjliga applikationer för offentlig sektor framförallt inom det polisiära arbetet, som kommer leda till gränsdragningsutmaningar och debatt.

- AI-system som innebär *hög risk* kan tillåtas om de genomgår en granskningsprocess av en behörig offentlig aktör och beviljas en CE-märkning. Denna ska bland annat innebära en adekvat nivå av transparens kring hur AI-systemet fungerar och dokumentation av systemets funktion för tillsyn och säkerställandet av användning av data som är av hög kvalitet. AI-rättsakten ställer också informations- och transparenskrav för hur AI-system som är avsedda att interagera med fysiska personer ska utformas och utvecklas. Det föreslås också att varje medlemsstat ska utse eller etablera minst en tillsynsmyndighet som ska ansvara för att säkerställa att de nödvändiga förfarandena följs. När förslaget publicerades presenterades också bedömningar av vilken typ av användningsområden som kunde ingå i högriskklassificeringen, och påfallande mycket av offentlig förvaltning fördes fram, varför föreslagna utvärderingar och kravställanden blir relevant för många av de beslutsprocesser som finns inom denna sektor.<sup>82</sup> Noterbart är även att kommissionen föreslås få rätt att genom delegerade akter vid behov lägga till ytterligare AI-system som ska definieras som högrisk, vilket gör att det finns en möjlig rörlighet över tid i vilka tillämpningar som träffas av regleringen.
- AI-system med *begränsad risk* kan ha transparenskrav – till exempel att en mänsklig användare ska veta att det är ett AI-system som interaktionen sker med – och system med minimal risk regleras inte enligt rättsakten.

Rättsakten anger straffavgifter, vars storlek beror på vilket brott mot förordningen som skett. Användningen av förbjudna AI-system ska leda till böter på upp till 30 000 000 euro eller upp till 6 procent av föregående

<sup>82</sup> <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>.

års globala omsättningen om det gäller ett företag vars omsättning är högre än det stadgade bötesbeloppet.

Vi kan konstatera att AI-rättsaktsförslaget är omfattande, varför också bedömningen av konsekvenserna alla dess komponenter i det här stadiet är svårt. Infrastrukturdepartementet skickade förslaget på remiss den 21 maj 2021, och bad om svar senast den 24 juni 2021 och även om man kan misstänka att många av remissinstanserna har haft svårt att bilda sig en initierad överblick på så kort tid så finns det några anmärkningar att särskilt notera.<sup>83</sup> Remissinstanser lyfter bland annat fram utmaningar med regelkonflikter eller otydlighet i relation till befintlig lagstiftning (DIGG, IMY). Vissa betonar övergångsproblematiken för redan etablerade AI-system (Arbetsförmedlingen), eller att det inte är tydligt hur skiljelinjen går mellan forskningens behov och användningen på den inre marknaden (bland annat Linköpings universitet och Formas).

#### *4.4.2 Definitionen i den föreslagna AI-rättsakten*

Flera av de svenska remissinstanser anmärker på den vida definitionen av artificiell intelligens i kommissionens förslag på AI-rättsakt (bland annat DIGG och Lunds universitet), vilket föranleder att ett särskilt fokus läggs vid den här. I både kommissionens första definition såväl som i AI-expertgruppens omformulerade version, som vi redogjort för ovan, finns ett mått av självständighet eller autonomi med som inte återfinns i förslaget på rättsakt. I huvudsak regleras AI-definitionen i rättsaktens Art. 3(1) som anger att AI-system är:

programvara som utvecklats med en eller flera av de tekniker och metoder som förtecknas i bilaga I och som, för en viss uppsättning människodefinierade mål, kan generera utdata såsom innehåll, förutsägelser, rekommendationer eller beslut som påverkar de miljöer som de samverkar med.

Denna definition ska läsas tillsammans med bilaga 1, som inkluderar flera begrepp av vid karaktär, exempelvis ”statistiska metoder” och ”optimeringsmetoder”:

- a) Metoder för maskininlärning, inbegripet övervakad, oövervakad och förstärkt inlärning, med hjälp av en mängd olika tillvägagångssätt, inklusive djupinlärning.

<sup>83</sup> Regeringskansliet ”Remiss av Europeiska kommissionens förslag till förordning om harmoniserade regler för artificiell intelligens”.

- b) Logik- och kunskapsbaserade metoder, inklusive kunskapsrepresentation, induktiv (logisk) programmering, kunskapsbaser, inferens- och deduktionsmotorer, (symboliska) resonemang och expertsystem.
- c) Statistiska metoder, bayesisk beräkning, sök- och optimeringsmetoder.

Utmaningen med en alltför vid definition påpekas bland annat av DIGG:s remissvar i termer av de kostnader som medföljer utvärderings- och uppföljningskraven. Vi har i kapitlets inledning konstaterat att ”AI” är ett mångbottnat och föränderligt koncept, vilket föranleder tydliga utmaningar när begreppsapparaten ska styra en reglering. Det finns en risk för att en alltför vid definition leder till att AI-rättsakten får karaktären av en mer allmän teknikreglering. Vi antar därför att definitionen kommer att bli mer strikt i lagstiftningens slutversion.

## 5 Övergripande slutsatser: Från etik till juridik

Som nämnt i inledningen fick DIGG, Arbetsförmedlingen, Bolagsverket, och Skatteverket i juni 2021 i uppdrag av regeringen att främja offentlig förvaltnings förmåga att använda artificiell intelligens (AI), i syfte att stärka Sveriges välfärd och konkurrenskraft.<sup>84</sup> Uppdraget hänvisade till behovet av en AI-guide, att anpassa efter internationella rekommendationer och riktlinjer, vilket vi kortfattat har redovisat ovan. Vi har även visat på hur vissa myndigheter redan har tagit sig an uppgiften och influerats av de idéer om behov av styrning som många av dessa riktlinjer uttrycker, inte minst gällande transparens, ansvarsfördelning och rättvisefrågor, ofta uttryckta i det något svåröversatta fairnessbegreppet.<sup>85</sup>

Genom att i det här kapitlet ge en överblick över hur AI används i svensk offentlig förvaltning och ställa den utvecklingen i kontrast till både de idéer och utmaningar som forskningen pekar på i den vidare principiella styrningsdiskussionen på europeisk nivå kan vi konstatera att:

- AI-regleringsidéerna har utvecklats mycket över relativt få år i en dialektik med insikter inom en flervetenskaplig forskning, och har lett till en flora av ”mjuk” reglering i form av rekommendationer och etiska

<sup>84</sup> Regeringen (21 juni 2021) ”Uppdrag att främja offentlig förvaltnings förmåga att använda artificiell intelligens”. Diarienummer: I2021/01825.

<sup>85</sup> Jämför exempelvis Larsson, S. (2018) ”Sjyst AI och normativ design”, i Akenine, D. & Stier, J. (red.) *Människor och AI*. Stockholm: AddAI.

riktlinjer. Förslaget på en europeisk AI-rättsakt pekar dock på att en formalisering i termer av en juridifiering av delar av dessa insikter är på gång, med betydande konsekvenser för både AI-utveckling och tillämpning av AI i offentlig förvaltning.

- Definitionen av AI är påfallande svårfångad, konceptuellt, och rörlig, forskningsmässigt, vilket leder till en specifik styrningsutmaning och heterogenitet i internationella rekommendationer såväl som i kommissionens förslag på AI-rättsakt. Detta är också en fråga för bedömningen av betydelsen av AI inom offentlig förvaltning – som finns i spännvidden mellan varianter av regelbaserad automation, som funnits länge, och mer autonomt lärande och prediktiva rekommendationssystem baserade på maskininläring, som är ett nyare fenomen.
- Flertalet av de svenska myndighetsrapporter som behandlar AI-relaterade frågor, och inte minst kommissionens förslag på AI-rättsakt, lägger tonvikt vid behovet av transparens i tillämpningen av AI-system. Begreppet är dock mångbottnat, och här betonas både behovet av att kunna granska och idka tillsyn över tillämpningar genom exempel på AI-register och förtroendemodellkort, såväl som både tjänstepersonens relation till AI-system – som ju kan ansvara för myndighetsbeslut där AI-system spelat en avgörande roll i en underliggande bedömning – och slutanvändarens förståelse för vem eller vad hen interagerar med.
- I litteraturen pekas bland annat på behovet av att bättre hantera utmaningar med transparens och ansvar i relationen mellan offentlig förvaltning och privat utförare. Här finns intressekonflikter mellan det allmännas behov av insyn och företagets ägarintresse och att exempelvis hemlighålla konkurrensfördelar. Här kan registerföring säkerställa nödvändig granskningsbarhet. Även upphandlingsområdet ser till exempel ut att spela en särskilt central roll för AI-utvecklingen inom offentlig förvaltning, både gällande granskningsbarhet såväl som för hur nyttan av offentligt insamlade data, som används för att träna upp kommersiella system, också säkerställs komma till det allmännas godo.
- Utvecklingen mot en rättslig formalisering kan vara gynnsam för utvecklingen av AI i offentlig sektor, eftersom bristen på klaggörande rättsregler kan vara ett hinder för myndigheter. Vi ser dock här ett behov av fler studier och klaggöranden kring hur den statliga värde-

grunden och principer för god förvaltning kan säkerställas även när AI-system används som beslutsstöd eller rentav beslutsfattare i offentlig förvaltning. Ytterst, för att fortsatt säkerställa ett upprätthållande av förvaltningens nödvändiga tillförlitlighet.

