

RÄTTSLIGT SKYDD MOT ALGORITMISK DISKRIMINERING I EUROPA:

Nuvarande ramverk och kvarvarande luckor



I samarbete med:

Interfederal Centre for Equal Opportunities (Unia), Belgien,
Diskrimineringsombudsmannen (YVV), Finland, och
Commission for Citizenship and Gender Equality (CIG), Portugal

Raphaële Xenidis

Co-funded
by the European Union



EUROPEAN UNION

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE

Co-funded and implemented
by the Council of Europe

RÄTTSLIGT SKYDD MOT ALGORITMISK DISKRIMINERING I EUROPA:

Nuvarande ramverk
och kvarvarande luckor

Raphaële Xenidis

Rapporten har tagits fram med ekonomiskt stöd från Europeiska unionen och Europarådet. Författaren är ensam ansvarig för innehållet i rapporten.

De åsikter som uttrycks i detta dokument kan inte på något sätt anses återspegla vare sig Europeiska unionens eller Europarådets officiella ståndpunkt.

Återgivning av utdrag (upp till 500 ord) är tillåten, förutom för kommersiella ändamål, under förutsättning att textens integritet bevaras, att utdraget inte används bortom sitt sammanhang, att det inte ger ofullständig information eller på annat sätt vilseleder läsaren om textens art, omfattning eller innehåll. Källtexten måste alltid anges på följande sätt: "© Europarådet, publiceringsåret". Alla andra förfrågningar om reproduktion/översättning av hela eller delar av dokumentet ska ställas till Publications and Visual Identity Division, Council of Europe (Europarådet, F-67075 Strasbourg Cedex eller publishing@coe.int).

All övrig korrespondens rörande detta dokument ska ställas till Hate Speech, Hate Crime and Artificial Intelligence Unit of the Council of Europe's Inclusion and Anti-discrimination Programmes Division, Council of Europe, F-67075 Strasbourg Cedex, Frankrike. E-post: anti-discrimination@coe.int

Den ursprungliga engelska texten är framtagen av Europarådet. Denna svenska översättning publiceras efter överenskommelse med Europarådet, men ansvaret för innehållet ligger helt och hållet hos översättarna. Europarådets avdelning för publikationer och visuell identitet har inte korrigerat några typografiska eller grammatiska fel i denna svenska publikation.

Omslagsdesign: Publications and Visual Identity Division (DPiV), Council of Europe
Layout av engelsk version: Jouve, Paris
Layout av svensk version: BGRAPHIC, Copenhagen
Översättning: ALM Translations
Juridisk granskning av översättningen:
Ada Möller, 1001 Lakes
Foton: Shutterstock

Innehåll

FÖRKORTNINGAR	5
ERKÄNNANDEN	6
INLEDNING	7
1. BEFINTLIGA RÄTTSLIGA RAMAR	9
1.1. Definitioner	9
1.2. Algoritmisk diskriminering i Europa	10
1.3. Användning av AI- och ADM-system inom offentliga förvaltningar och sektorer av intresse	11
1.4. Befintliga rättsliga ramar	15
1.4.1. Styrning av AI- och ADM-system	15
1.4.2. Instrument för att motverka diskriminering	28
1.4.3. Instrument för dataskydd	33
1.4.4. Rättspraxis från Europeiska domstolen för de mänskliga rättigheterna och Europeiska unionens domstol	36
1.5. Viktiga intressenter	39
2. RÄTTSLIGA BRISTER	43
2.1. Motivering för införande av AI- och ADM-system	43
2.2. Bristande transparens	44
2.3. Frågor om tillgång till rättslig prövning	46
2.4. Tillämpning av befintliga och framtida bestämmelser	48
2.4.1. Institutionellt samarbete	48
2.4.2. Konceptuella luckor	48
2.4.3. Begränsningar av insamling och behandling av data	50
2.5. "Riskbegränsa" AI-system	50
2.6. Tekniska standarder, likabehandlingsorgan och frågan om harmonisering	51
SAMMANFATTNING – ANVÄNDNING AV AI I BELGISK OFFENTLIG FÖRVALTNING OCH POLICYREKOMMENDATIONER OM ALGORITMISK DISKRIMINERING	55
SAMMANFATTNING – ANVÄNDNING AV AI I FINSK OFFENTLIG FÖRVALTNING OCH POLICYREKOMMENDATIONER OM ALGORITMISK DISKRIMINERING	59
SAMMANFATTNING – ANVÄNDNING AV AI I PORTUGISISK OFFENTLIG FÖRVALTNING OCH POLICYREKOMMENDATIONER OM ALGORITMISK DISKRIMINERING	63
REFERENSER	67

Förkortningar

AI	Artificiell intelligens
ADM	Automatiserat beslutsfattande
CSOs	Civilsamhällesorganisationer
CAI	Europarådets kommitté för artificiell intelligens
CEN	Europeiska standardiseringskommittén
CENELEC	Europeiska kommittén för elektroteknisk standardisering
FRIA	Konsekvensbedömning avseende grundläggande rättigheter
GDPR	Allmänna dataskyddsförordningen
GPAI	Uppförandekoden för AI för allmänna ändamål
HUDERIA	Risk- och konsekvensbedömning av AI-system med hänsyn till mänskliga rättigheter, demokrati och rättsstatsprincipen
TEU	Fördraget om Europeiska unionen
TFEU	Fördraget om Europeiska unionens funktionssätt

Erkännanden

Denna rapport har utarbetats inom ramen för projektet "Främjandet av jämlikhet och icke-diskriminering av Likabehandlingsorgan i samband med användning av artificiell intelligens inom den offentlig sektorn", som finansieras av Europeiska unionen via instrumentet för tekniskt stöd och samfinansieras av Europarådet. Projektet genomförs av Europarådet i samarbete med Europeiska kommissionen, Interfederal Centre for Equal Opportunities (Interfederala centret för lika möjligheter, Unia, Belgien), Diskrimineringsombudsmannen (Finland) och Commission for Citizenship and Gender Equality (Kommissionen för medborgarskap och jämställdhet, Portugal).

Rapporten har författats av Raphaële Xenidis (biträdande professor i europeisk rätt, Sciences Po Law School, Frankrike). Sammanfattningarna av de nationella sammanhangen har hämtats från Ine van Zeelands (Europarådets nationella konsult) arbete med det belgiska avsnittet, Emeline Banzuzis (Europarådets nationella konsult) arbete med det finska avsnittet och Eduardo dos Santos (Europarådets nationella konsult) arbete med det portugisiska avsnittet.

Europarådet vill tacka projektets mottagande institutioner och Europeiska kommissionen för deras ihållande engagemang under hela arbetsprocessen, och i synnerhet till: Nele Roekens, Nadine Brauns och Thelma Raes (Interfederal Centre for Equal Opportunities, Unia, Belgien); Tiina Valonen och Ville Rantala (Diskrimineringsombudsmannen, YVV, Finland); Carla Peixe, Ana Martinho Fernandes, Alexandra Andrade och Susana Miguel (Commission for Citizenship and Gender Equality, CIG, Portugal), och Massimiliano Santini (Task Force for Reform and Investments [Arbetsgruppen för reformer och investeringar], Secretariat-General, Europeiska kommissionen) samt Menno Ettema, Sara Haapalainen, Ayça Dibekoğlu och Delfine Gaillard (Anti-discrimination Department [Avdelningen för antidiskriminering], Europarådet).

Rapporten har också gynnats av bidrag från nationella experter från offentlig och privat sektor i Belgien, Finland och Portugal som deltog i intervjuer i slutet av 2024 och början av 2025. Deras perspektiv berikade de nationella avsnitten i rapporten.





Inledning

Den ökande integrationen av artificiell intelligens (AI) inom olika områden och sektorer innebär stora utmaningar för skyddet av de grundläggande rättigheterna. Bland dessa framstår algoritmisk diskriminering som ett särskilt angeläget problem. Empirisk forskning har visat att algoritmisk bias inte bara återspeglar utan också förvärrar befintliga sociala ojämlikheter i stor skala, särskilt inom områden där algoritmiskt beslutsfattande är vanligt förekommande, som polisarbete, rekrytering, utbildning och försäkringar. Under en längre period utgjorde skärningspunkten mellan icke-diskriminering och dataskyddslagstiftning den primära rättsliga grunden för att hantera fall av algoritmisk diskriminering i ett europeiskt sammanhang.

Den senaste tidens rättsliga, institutionella och politiska utveckling har dock på ett betydande sätt förändrat styrningslandskapet för AI i Europa. År 2024 antog både Europarådet och Europeiska unionen viktiga rättsliga instrument: Europarådets Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (ramkonventionen om artificiell intelligens och mänskliga rättigheter, demokrati och rättsstatsprincipen [ramkonventionen om AI]) och Europaparlamentets och rådets förordning (EU) 2024/1689 från den 13 juni 2024 om fastställande av harmoniserade regler för artificiell intelligens (EU:s AI-förordning). Dessa instrument fastställer omfattande styrningsramverk som syftar till att skydda grundläggande rättigheter – inklusive rätten till icke-diskriminering – i samband med AI och system

för automatiserat beslutsfattande (ADM). Antagandet av dessa ramverk förväntas få betydande rättsliga och institutionella konsekvenser på nationell nivå, särskilt med tanke på att medlemsländerna senast 2026 kommer att vara skyldiga att anpassa sin nationella lagstiftning och sina administrativa strukturer till de skyldigheter som anges i EU:s AI-förordning.

AI-förordningen understryker nödvändigheten av att AI-teknik och dess styrande regelverk utvecklas i enlighet med EU:s värderingar, såsom de fastställs i artikel 2 i fördraget om Europeiska unionen (TEU), de grundläggande rättigheter och friheter som garanteras i fördragen och stadgan om de grundläggande rättigheterna i enlighet med artikel 6 i TEU. Det står uttryckligen att AI måste vara en människocentrerad teknologi. På samma sätt gör Europarådets ramkonvention om AI gällande att man måste säkerställa att alla aktiviteter som ingår i AI-livscykeln är helt förenliga med mänskliga rättigheter, demokrati och rättsstatsprincipen.

Denna rapport undersöker hur dessa nya rättsliga instrument bidrar till att stärka skyddet mot algoritmisk diskriminering i Europa och bedömer de luckor som fortfarande påverkar dessa rättsliga ramar.

Samtidigt syftar två nya EU-direktiv om likabehandlingsorgan till att stärka dessa institutioners roll och kapacitet i hela Europa (Europeiska unionen 2024a, 2024b). Likabehandlingsorgan, tillsammans med nationella institutioner för mänskliga rättigheter och ombudspersoner, förväntas spela en central roll i styrningen av AI- och ADM-system, särskilt för att säkerställa efterlevnaden av normer för grundläggande rättigheter, inklusive principen om icke-diskriminering. Denna rapport syftar därför till att spåra och sammanfatta utvecklingen för att förse relevanta aktörer, såsom likabehandlingsorgan, med verktyg för att anpassa sig till ett föränderligt rättsligt landskap och, där så är relevant, ingripa i den pågående utvecklingen.

Rapporten är uppbyggd i två huvuddelar:

- en bedömning av befintliga rättsliga ramar
- en utvärdering av kvarvarande luckor.

I det första avsnittet ges definitioner (1) och exempel på användning av AI- och ADM-system i offentlig sektor och i andra sektorer av intresse på europeisk nivå (2). Efter att ha förklarat hur algoritmisk diskriminering uppstår (3) kartläggs de nuvarande rättsliga, processuella och styrningsrelaterade ramarna för anti-diskriminering, jämlikhet och AI i Europa (4). Avsnittet avslutas med en undersökning av relevanta intressenters nuvarande och potentiella roll i samband med algoritmisk diskriminering (5).

I det andra avsnittet kartläggs rättsliga luckor i förhållande till dessa ramar och deras tillämpning. Denna rapport omfattar i första hand AI- och ADM-system som används av offentlig sektor i Europa och i andra hand system som används inom den privata sektorn, t.ex. stora bemanningsföretag.



1. Befintliga rättsliga ramar

1.1. Definitioner

I den här rapporten används begreppen "AI" och "ADM-system" för att täcka verksamheten i inlärningsbaserade och regelbaserade system, som både kan vidmakthålla och förstärka diskriminering. I artikel 2 i Europarådets ramkonvention om AI definieras ett AI-system som "ett maskinbaserat system som, för uttryckliga eller underförstådda mål, drar slutsatser härledda från den indata det tar emot om hur utdata såsom förutsägelser, innehåll, rekommendationer eller beslut som kan påverka fysiska eller virtuella miljöer ska genereras". Det medger att "[de o]lika AI-system varierar i fråga om autonomi och anpassningsförmåga efter införandet" (Europarådet 2024a). Denna definition motsvarar den definition som ges i artikel 3.1 i EU:s AI-förordning.¹ Inom AI-teknikerna har i synnerhet maskininläring kommit att användas i stor utsträckning för att identifiera så kallade korrelationer mellan datapunkter

1. I Europaparlamentets och rådets förordning (EU) 2024/1689 av den 13 juni 2024 som fastställer harmoniserade regler om artificiell intelligens, definierar artikel 3.1 ett AI-system som "ett maskinbaserat system som är utformat för att fungera med varierande grad av autonomi och som kan uppvisa anpassningsförmåga efter införande och som, för uttryckliga eller underförstådda mål, drar slutsatser härledda från den indata det tar emot, om hur utdata såsom förutsägelser, innehåll, rekommendationer eller beslut som kan påverka fysiska eller virtuella miljöer ska genereras".

i syfte att göra förutsägelser och rekommendationer eller poängsätta, klassificera och rangordna föremål eller personer.² Vissa ADM-system saknar däremot den grad av anpassningsförmåga och autonomi som krävs för att kvalificera sig som AI-system, men är ändå viktiga att undersöka i denna rapport med tanke på deras diskriminerande potential. Regelbaserade system kan till exempel felklassificera individer och automatisera beslutsprocesser utifrån detta, med potentiellt diskriminerande konsekvenser. Begreppet "ADM-system" omfattar ett brett spektrum av beslutssystem, från helt automatiserade system som inte kräver någon mänsklig kontroll, enligt definitionen i t.ex. artikel 22 i EU:s allmänna dataskyddsförordning (GDPR), till halvautomatiserade system eller beslutsfrämjande system som omfattar både maskinstyrd och mänsklig kontroll och som används i mycket stor utsträckning i praktiken.³ Denna rapport fokuserar både på AI- och ADM-system.

1.2. Algoritmisk diskriminering i Europa

Diskriminering i algoritmiskt beslutsfattande uppstår när automatiserade system på ett orättvist sätt missgynnar vissa individer eller grupper. Det finns flera viktiga faktorer som bidrar till algoritmisk diskriminering.

För det första kan bias i utbildnings- och utvärderingsdata leda till diskriminerande beslut. Det är inte bara så att skev datainsamling och dataproduktion kan leda till icke-representativa datauppsättningar, utan när AI och datadrivna ADM-system tränas på eller bearbetar historiska data som återspeglar fördomar mot, ojämlikhet mellan eller uteslutning av till exempel vissa etniska, könsrelaterade eller socioekonomiska grupper, har de också en tendens att reproducera och/eller förstärka dessa fördomar. Till exempel kan en urvalsalgoritm som tränas på historiska anställningsdata orättvist gynna manliga kandidater genom att reproducera tidigare anställningsbeslut som var snedvridna av diskriminerande stereotyper mot kvinnor eller minoritetsgrupper.

För det andra kan de designval som görs av utvecklarna, inklusive valet av målfunktioner och prediktorer, typen av modell, rättvisemått och tröskelvärden, återspegla diskriminerande fördomar. Om utvecklare inte tar hänsyn till systematiska ojämlikheter kan AI- och ADM-system oavsiktligt fatta beslut som oproportionerligt skadar vissa grupper. Till exempel använder välfärdstjänster i Europa alltmer ADM-system för att förutsäga bedrägerier bland mottagare av sociala förmåner. Det har visat sig att själva kategorin "risk" och de prediktorer som används i dessa ADM-system i verkligheten återspeglar stereotyper gentemot individer som lever i fattigdom, stereotyper som är djupt rotade i välfärdstjänsternas praxis, strukturer och institutioner (Dubois 2021). Detta problem återspeglar den utbredda s.k. proxydiskrimineringen,

2. Enligt Binns är "[d]e flesta användningar av maskininlärning inom den offentliga sektorn av den övervakade typen. Med övervakning avses att inlärningsalgoritmen måste få veta vad en beslutsfattare vill förutsäga eller klassificera, till skillnad från oövervakade metoder som är utformade för att upptäcka latenta strukturer i en datauppsättning". Se Binns R. (2020), "Algorithmic decision-making: a guide for lawyers", *Judicial Review*, vol. 25, nr 2, s. 3-4.

3. För en taxonomi av dessa olika ADM-system, se Palmiotto F. (2024), "When is a decision automated? A taxonomy for a fundamental rights analysis", *German Law Journal*, vol. 25, nr 2, s. 210-36, tillgänglig på: <https://doi.org/10.1017/glj.2023.112>, besökt den 6 november 2025.

dvs. diskriminering som baseras på algoritmisk bearbetning av till synes neutrala proxyvariabler som osynligt kodar ojämlikheter.

För det tredje är det svårt att övervaka eller ifrågasätta algoritmisk diskriminering på grund av bristen på transparens och ansvarsskyldighet för de AI- och ADM-system som för närvarande används. Många AI- och ADM-system, särskilt de som bygger på komplexa modeller för maskininlärning, fungerar som "svarta lådor", vilket innebär att deras beslutsprocesser inte är transparenta. Denna brist på transparens gör det svårt att identifiera när och hur diskriminering förekommer, vilket förhindrar effektiv granskning och ansvarsskyldighet. Till exempel visade skandalen i Nederländerna 2021, som orsakades av en algoritm som användes för att förutse bedrägerier bland mottagare av barnomsorgsbidrag, hur svårt det var för de felaktigt anklagade att förstå hur och varför de hade klassificerats som bedragare, än mindre att ifrågasätta beslutet (Amnesty International 2021).

Otillräcklig testning och övervakning från leverantörer och användare av AI- och ADM-system, både inom den offentliga och den privata sektorn, ökar dessutom sannolikheten för att diskriminering uppstår utan förvarning. Test- och övervakningsprocesser måste finnas på plats före, under och efter användning. Genom att involvera relevanta civilsamhällesorganisationer, likabehandlingsorgan och institutioner för mänskliga rättigheter går det att säkerställa att dessa förfaranden är effektiva.

Hastigheten och omfattningen av AI- och ADM-system riskerar slutligen att sprida diskriminering på en systemisk nivå. Branschens löfte om effektivitet och kostnadsbesparingar har uppmuntrat många organisationer, inklusive offentliga förvaltningar, att rationalisera beslutsfattandet genom att använda AI- och ADM-system. Dock utan att investera tillräckliga resurser i att förebygga och mildra algoritmisk diskriminering, till exempel genom att utbilda handläggare och investera i representativa data, lämpliga revisions- och testmekanismer och ansvarsprocesser, kommer AI- och ADM-system att vara diskriminerande. Branschens berättelse, enligt vilken AI- och ADM-system möjliggör en drastisk minskning av kostnaderna, måste därför ifrågasättas. Det kan mycket väl vara så att det är kostsamt att genom lämpliga skyddsåtgärder och investeringar se till att AI- och ADM-system fungerar lagligt, det vill säga utan att orsaka diskriminering.

1.3. Användning av AI- och ADM-system inom offentliga förvaltningar och sektorer av intresse

AI och ADM-system används i allt större utsträckning i både den privata och den offentliga sektorn i Europa. En genomgång av litteraturen visar exempel på tillämpningar inom många områden, t.ex. hälso- och sjukvårdssektorn, finans- och banksektorn, transportsektorn och personalområdet. Offentliga förvaltningar inför eller experimenterar gradvis med AI- och ADM-system inom områden som skattefrågor, transport, social trygghet, migration samt rättsväsende och polis. Applikationerna kan utvecklas internt eller köpas in från externa privata företag. Förutom godartade administrativa stöduppgifter används AI- och ADM-system inom dessa områden främst för att upptäcka och övervaka bedrägerier, vilket medför stora risker för diskriminering. Andra vanliga tillämpningar är chatbottar och virtuella assistenter, som också kan uppvisa diskriminerande drag men troligen med mindre

skadliga effekter. Med detta sagt verkar användningen av AI respektive ADM-system att skilja sig åt: medan AI-applikationer som utvecklats internt för närvarande ofta befinner sig i experimentstadiet, verkar ADM-system användas i större utsträckning.

Utmaningar i samband med kartläggning av användningen av AI- och ADM-system

Inventeringsarbetet som genomförs i detta avsnitt kommer med en viktig varning: att kartlägga offentliga förvaltningars användning av AI- och ADM-system är en svår uppgift. Forskning lyfter fram bristen på tydlig och konsekvent information om hur offentliga förvaltningar använder AI- och ADM-system. I vissa länder, som Finland, har offentliga förvaltningar informationsskyldighet och måste publicera information om användningen av ADM-system på sina webbplatser samt informera berörda personer om sådan användning när de fattar beslut. Databaser som systematiskt registrerar AI- och ADM-tillämpningar är dock fortfarande underutvecklade.⁴ Relevanta intressenter som likabehandlingsorgan, nationella institutioner för mänskliga rättigheter och ombudspersoner har därför svårt att utöva sitt mandat att utreda, övervaka, identifiera och ifrågasätta fall av algoritmisk diskriminering samt att stödja offer. Dessa hinder äventyrar en effektiv tillämpning av rätten till icke-diskriminering. Andra strategier som används för att få information (t.ex. förfrågningar om informationsfrihet, parlamentariska utredningar) stöter också på utmaningar och begränsningar och ger inte alltid tillräcklig insyn i hur offentliga förvaltningar använder AI- och ADM-system. I Frankrike, till exempel, resulterade en begäran enligt informationsfriheten från CSO:er för att få klarhet i ett riskbedömningssystem som används av socialförvaltningen som hanterar familjebidrag (CNAF) i att man offentliggjorde en begränsad del av källkoden för tidigare modeller (i motsats till den nuvarande modellen) med reviderade listor över använda variabler (La Quadrature du Net 2023). Andra utmaningar är bristen på insyn i användningen av AI- och ADM-system, vilket i sig föregriper frågor och förfrågningar: om offentliga förvaltningar inte rapporterar om användningen av sådana system blir det svårt att veta vart man ska rikta sina förfrågningar.

Rekrytering

Rapporterade användningsområden inkluderar profileringsverktyg som används av offentliga arbetsförmedlingstjänster för att förutspå chanser till anställning eller arbetslöshet, vilket används för att hjälpa handläggare att fördela stödresurser till arbetssökande. Vissa av dessa profileringsverktyg har visat sig vara diskriminerande. I Österrike visade det sig till exempel att en prototyp som utvecklats av den österrikiska arbetsförmedlingen, kallad AMS-algoritmen, som syftade till att förutspå arbetssökandes utsikter till anställning för att hjälpa handläggarna att besluta om resursfördelning, ledde till diskriminering av bland annat ensamstående mödrar och arbetssökande med migrationsbakgrund (Allhutter et al. 2020). Liknande system har sedan 2018 använts av den franska arbetsförmedlingen France Travail för att bedöma risken för bedrägerier, förutse avhopp och bedöma deltagarnas anställningsbarhet (La Quadrature du Net 2024). Liknande risker har också uppmärksammats för

4. Detta kommer troligen att ändras efter implementeringen av AI-förordningen.

diskriminering av individer i socialt och ekonomiskt utsatta situationer, men också på grundval av kön, etnicitet och funktionshinder. Andra problem är bristen på adekvat vägledning och utbildning för handläggare som använder AI- och ADM-verktyg, samt de sätt på vilka arbetssökandes personuppgifter används i profileringsprogram. De offentliga arbetsförmedlingarna kan också använda andra typer av verktyg, som till exempel matchningsalgoritmer, för att rekommendera lediga jobb till arbetssökande.

Brottsbekämpning

AI- och ADM-system används också regelbundet av brottsbekämpande myndigheter. Forskning visar dock att teknik för ansiktsigenkänning kan ha diskriminerande bias.⁵ Även när rapporter visar att AI- och ADM-system i sig inte är partiska kritiserar distribueringen av dem för att på ett oproportionerligt sätt rikta in sig på och övervaka minoriteter och för att utföra etnisk profilering.⁶ Ofta är bevisen för sådan användning fortfarande indicier, och bristen på transparent information om vilka system som används och hur de används förhindrar systematiska undersökningar av de risker som uppstår för grundläggande rättigheter, inklusive icke-diskriminering. Som framgår nedan (se avsnitt 1.4.1.) medger AI-förordningen omfattande undantag för polisen och brottsbekämpande myndigheter när det gäller användningen av igenkänningsteknik som behandlar bilder, till exempel ansikten och andra biometrisk datauppgifter.

Socialomsorg och välfärdssektorer

Olika delar av välfärds- och socialförsäkringssystemen i Europa experimenterar för närvarande med, eller lanserar, AI- och ADM-system (t.ex. Österrike, Belgien, Danmark, Frankrike, Polen, Portugal, Slovenien, Sverige, Nederländerna). I Frankrike syftar systemet som används av den nationella myndighet som ansvarar för fördelningen av familjebidrag till att förutse risker för bedrägerier och fel bland mottagarna för att hjälpa handläggarna att rikta in sina kontroller (La Quadrature du Net 2023). Riskerna för diskriminering har uppmärksamats av CSO:er, och ett fall har tagits upp i franska Conseil d'État där man bland annat pekar på diskriminering på grund av kön, familjestatus, ålder och funktionshinder (se Conseil d'État 2024). I Nederländerna ledde ett liknande system till att regeringen avgick 2021 efter att det visat sig att

5. CSO Liberty kritiserade ett program för ansiktsigenkänning som användes av South Wales Police bland annat för diskriminering på grund av kön och/eller etnicitet eftersom det gav en högre andel positiva träffar för kvinnoansikten och/eller för ansikten tillhörande svarta och etniska minoriteter. Se det efterföljande beslutet från UK Court of Appeal i *R (Bridges) v. Chief Constable of South Wales Police* ([2020] EWCA Civ 1058), som framhöll att South Wales Police "konsekvensbedömning avseende lika behandling var uppenbart otillräcklig och grundade sig på ett rättsligt fel (genom att inte erkänna risken för indirekt diskriminering)" och att dess "efterföljande tillvägagångssätt för att bedöma eventuell indirekt diskriminering till följd av användningen av AFR är bristfälligt".

6. Detta sker till exempel när polismyndigheter sparar bilder av felaktiga matchningar under en oproportionerligt lång tid i sina system, eller när överbevakning av platser som till stor del bebos av rasifierade grupper återkopplas till prediktiva verktyg och utökar massövervakningen av dessa grupper.

mottagarna diskriminerats på grund av ras⁷, etniskt ursprung och medborgarskap.⁸ Inom utbildningsområdet använde den nederländska utbildningsmyndigheten en riskprofileringsalgoritm för att upptäcka bedrägerier i samband med mottagande av studiebidrag (Algorithm Audit 2024a). Myndigheten erkände att "[s]tuderter med utomeuropeisk migrationsbakgrund tilldelades en högre riskpoäng via en riskprofil och valdes oftare ut för hembesök", vilket skapade indirekt diskriminering (Algorithm Audit 2024b). I Belgien upphörde OASIS-systemet 2023 efter att forskare påvisat risker för diskriminering på grund av fattigdom. Sådana system kombinerar uppgifter från olika förvaltningar (t.ex. om skatt, anställning, familjeförmåner, pensioner), ibland i kombination med uppgifter som samlas in av privata företag (t.ex. om energiförbrukning), för att profilera användare och fastställa riskpoäng.

Migration och medborgarskap

AI- och ADM-system används också av vissa offentliga förvaltningar i Europa för stöd i beslutsfattande inom migrationsområdet, till exempel när det gäller beslut om medborgarskap, asyl eller uppehållstillstånd. Rapporterade användningsområden är till exempel identifiering och bedömning av språk, upptäckt av bedrägerier som rör identitetshandlingar, ärendehantering, interaktion med migranter, bland annat genom chattbotar, migrationsprognoser och teknik för gränsövervakning (European Migration Network and Organisation for Economic Co-operation and Development 2022; European Network of National Human Rights Institutions 2024; McGregor and Molnar 2023). Även om dessa tekniker påstås användas för att effektivisera migrationshanteringen kan de ha en negativ inverkan på migranternas rättigheter genom att förstärka diskriminering, skapa oro för integritets- och dataskydd och avskräcka migranter från att söka skydd (European Network of National Human Rights Institutions 2014). Enligt FN:s särskilda rapportör om samtida former av rasism, rasdiskriminering, främlingsfientlighet och liknande intolerans omfattar "digitala gränser" en rad infrastrukturer för styrning som "i allt högre grad bygger på maskin-inläring, big data, automatiserade algoritmiska beslutssystem, prediktiv analys och relaterad digital teknik" (FN:s särskilda rapportör 2021). Den teknik som används omfattar "identifieringssystem, system för ansiktsgenkänning, marksensorer, drönare för videoövervakning, biometrisk databaser och till och med beslutsprocesser för visum och asyl" (ibid.).

Skattemyndigheter

Även om det generellt sett finns mindre information om skattemyndigheternas användning av AI- och ADM-system, experimenterar flera europeiska länder för närvarande med sådana tekniker, främst i syfte att upptäcka bedrägerier. Till exempel används AI- och ADM-system för att flagga misstänkta filer för vidare utredning. Myndigheterna kan vara ovilliga att lämna ut information om dessa system av rädsla

7. Eftersom alla människor tillhör samma art avvisar Europarådets ministerkommitté teorier som bygger på förekomsten av olika "raser". I detta dokument används dock termen "ras" för att säkerställa att personer som allmänt och felaktigt uppfattas som "tillhörande en annan ras" inte utesluts från det skydd som lagstiftningen och genomförandet av politiken ger.

8. SyRI-systemet ansågs också ha inneburit ett oproportionerligt intrång i slutanvändarnas rätt till privatliv eftersom det behandlade personuppgifter från olika statliga myndigheter.

för att användarna ska "lura" systemet. Det finns risk för diskriminering eftersom system för upptäckt av bedrägerier kan påverka personer oproportionerligt på grund av skyddade egenskaper, som till exempel socioekonomisk status, migrationsbakgrund eller geografisk plats, även om dessa kategorier inte uttryckligen används som riskfaktorer i dessa system. I det ovan nämnda SyRI-ärendet i Nederländerna misstänkte skattemyndigheterna 26 000 familjer för bedrägeri med sociala bidrag. Algoritmen flaggade oproportionerligt etniska minoriteter och icke-nederländska medborgare.

Privat sektor

Inom den privata sektorn används AI- och ADM-system av företag som erbjuder personaltjänster. Exempel på tillämpningsområden är screening och profilering av arbetssökande, matchning av arbetsansökningar med deras profiler eller att utarbeta och översätta lediga tjänster. Dessa tillämpningar kan i olika grad innebära risker för diskriminering. På samma sätt används AI- och ADM-system i stor utsträckning inom finans och handel, till exempel marknadsföring, kundservice (t.ex. chattbottar), dynamisk prissättning och transaktionsövervakning. I Finland nekades till exempel en kund en onlinekredit av ett finansbolag som använde ett poängsystem baserat på uppgifter som kön, språk, ålder och bostadsort för att bedöma risken för att lånet inte skulle återbetalas (Algorithm Watch 2018). Diskriminerings- och jämställdhetsnämnden i Finland (2017) drog slutsatsen att detta utgjorde direkt diskriminering på flera skyddade grunder.

1.4. Befintliga rättsliga ramar

I detta avsnitt analyserar vi befintliga förordningar, lagar, rättspraxis, förfaranden, policyer, institutioner och så vidare, och undersöker hur de hanterar diskriminering som uppstår till följd av användningen av AI- och ADM-system.

1.4.1. Styrning av AI- och ADM-system

Europarådet

På Europarådsnivå finns det flera instrument som behandlar frågan om algoritmisk diskriminering eller som kan användas för att göra det.

För det första syftar Europarådets ramkonvention om AI, som antogs 2024 men ännu inte har ratificerats för att träda i kraft, till att säkerställa att aktiviteter inom systemprocesser för artificiell intelligens är helt förenliga med mänskliga rättigheter, demokrati och rättsstatsprincipen. I ingressen till ramkonventionen om AI lyfter man att AI kan "främja ... jämställdhet och egenmakt för alla kvinnor och flickor". Samtidigt uttrycks konventionsparternas "riskerna för diskriminering i digitala sammanhang, särskilt sådana som inbegriper AI-system, och dessas potentiella effekt att skapa eller förvärra ojämlikheter, bland annat sådana som drabbar kvinnor och personer i utsatta situationer när det gäller deras åtnjutande av sina mänskliga rättigheter och deras fulla, jämlika och faktiska deltagande i ekonomiska, sociala, kulturella och politiska frågor".

I artikel 10 om jämlikhet och icke-diskriminering i ramkonventionen om AI åläggs konventionsstaterna att "anta eller upprätthålla åtgärder i syfte att säkerställa att verksamhet under AI-systemens livscykel⁹ bedrivs med respekt för jämlikhet, inbegripet jämställdhet, och förbudet mot diskriminering, i enlighet med tillämplig internationell och inhemsk rätt". Konventionsparterna måste också "anta eller upprätthålla åtgärder som syftar till att övervinna ojämlikheter för att uppnå skäliga, riktiga och rättvisa resultat, i enlighet med dess tillämpliga inhemska och internationella skyldigheter i fråga om mänskliga rättigheter, när det gäller verksamhet under AI-systemens livscykel". Dessa skyldigheter kräver att medlemsländerna ser över och, om nödvändigt, reformerar sin lagstiftning för att säkerställa att icke-diskrimineringslagstiftningen fångar upp algoritmisk diskriminering. Såsom bekräftas i punkt 77 i motiveringen innehåller konventionen också en skyldighet för staterna att vidta åtgärder för att övervinna strukturella och historiska ojämlikheter i samband med aktiviteter inom AI-systemens livscykel. I punkterna 72–73 i motiveringen bekräftas att förbudet mot algoritmisk diskriminering i konventionen bygger på ett människorättsbaserat förhållningssätt till algoritmisk diskriminering som integrerar globala och regionala ramverk för mänskliga rättigheter. I punkterna 75–76 i motiveringen finns en förteckning över välkända källor till och typer av algoritmiska biaser. Bland "de olika sätt på vilka bias avsiktligt eller oavsiktligt kan införlivas i system för artificiell intelligens i olika skeden av deras livscykel", citerar motiveringen följande:

- ▶ Utveckling ("på grund av utvecklarnas medvetna eller omedvetna stereotyper eller bias"), modellering ("potentiella biaser som är inbyggda i den modell som systemen bygger på").
- ▶ Data (felaktiga eller otillräckligt representativa datauppsättningar i utbildnings-, aggregerings- eller utvärderingsfasen).
- ▶ Driftsättning ("bias som uppstår när sådana system implementeras i verkliga miljöer").
- ▶ Tolkning ("automatiserings- eller bekräftelsebias, varigenom människor kan sätta omotiverad tillit till maskiner och tekniska artefakter eller situationer där de väljer information som stöder deras egna åsikter ... ignorerar sin egen potentiellt motsägelsefulla bedömning och validerar algoritmiska resultat utan att ifrågasätta dem").
- ▶ Teknisk bias ("uppstår på grund av problem vid tillämpning av maskininlärning som resulterar i ytterligare bias som inte finns i de data som används för att träna systemet eller fatta beslut").
- ▶ Social bias ("underlåtenhet att på ett korrekt sätt ta hänsyn till historiska eller nuvarande ojämlikheter i samhället i aktiviteter inom livscykeln avseende systemprocesser för artificiell intelligens").

9. Se punkterna 14–15 i motiveringen till ramkonventionen: "Denna hänvisning till livscykeln säkerställer en övergripande strategi för att hantera AI-relaterade risker och negativa effekter på mänskliga rättigheter, demokrati och rättsstatsprincipen genom att fånga upp alla stadier av aktiviteter som är relevanta för system för artificiell intelligens." Exempel på relevant verksamhet är bland annat: "(1) planering och utformning, (2) insamling och bearbetning av data, (3) utveckling av system för artificiell intelligens, inklusive modellbygge och/eller finjustering av befintliga modeller för specifika uppgifter, (4) testning, verifiering och validering, (5) leverans/tillhandahållande av de tillgängliga systemen, (6) driftsättning, (7) drift och övervakning, och (8) avveckling."

För att underlätta verkställigheten av de skyldigheter som ingår i konventionen har Europarådet utformat en "metod [för att] säkerställa ett enhetligt tillvägagångssätt för identifiering, analys och utvärdering av risker och konsekvensbedömning av [AI-system] i förhållande till tillgodoseendet av de mänskliga rättigheterna, demokratis funktion och rättsstatens principer" som kallas HUDERIA (Europarådet 2022a). Den består av en sammanhangsbaserad riskanalys, en process för att engagera intressenter, en risk- och konsekvensbedömning och en plan för riskreducering, och kräver iterativ översyn. Även om tillvägagångssättet skiljer sig från den riskbaserade klassificering som antagits enligt AI-förordningen (se avsnittet om Europeiska unionen nedan), innehåller HUDERIA-metoden liknande inslag genom ett "graderat och differentierat tillvägagångssätt avseende åtgärder för identifiering, bedömning, förebyggande och begränsning av risker och konsekvenser som tar hänsyn till hur allvarliga och sannolika de negativa konsekvenserna är för mänskliga rättigheter, demokrati och rättsstatsprincipen samt relevanta kontextuella faktorer" (Europarådet 2024b). Vid fastställandet av risknivåer är det till exempel avgörande att ta hänsyn till de potentiella skadornas omfattning, allvarlighetsgrad, reversibilitet och sannolikhet. "Nollfrågor", dvs. frågor om huruvida AI- och ADM-system är relevanta och lämpliga för att utföra vissa uppgifter, måste också beaktas, med den potentiella konsekvensen att beslut om att inte utveckla eller använda sådana system kan fattas om riskerna för kränkningar av de mänskliga rättigheterna överväger de potentiella fördelarna. HUDERIA-metoden för risk- och konsekvensbedömning syftar till att säkerställa kompatibilitet "med de befintliga efterlevnadsrutiner som följs av branschen" (Europarådet 2022a) och kan främja offentliga och privata aktörers förebyggande av algoritmisk diskriminering. Även om det i artikel 16 i konventionen föreskrivs att ett ramverk för risk- och konsekvenshantering ska inrättas, är det inte obligatoriskt att följa HUDERIA för att uppfylla skyldigheterna enligt konventionen. Även om HUDERIA inte är ett rättsligt bindande instrument och inte ger någon tolkningsvägledning i förhållande till konventionen, kommer den att utgöra ett särskilt användbart verktyg för att identifiera, bedöma och minska riskerna för algoritmisk diskriminering.

Dessutom har olika rekommendationer publicerats för att kräva att konventionsstaterna tar itu med algoritmisk diskriminering i sina rättsliga och processuella ramverk. Till exempel uppmanade den parlamentariska församlingen i sin resolution 2343 (2020) om "förebyggande av diskriminering som uppstår från användningen av artificiell intelligens" medlemsstaterna att göra följande:

Se över sin antidiskrimineringslagstiftning och ändra den vid behov, för att säkerställa att den omfattar alla fall där direkt eller indirekt diskriminering, inklusive diskriminering på grund av association, kan orsakas som en följd av att AI använts, och att målsägande har full tillgång till rättslig prövning. I det senare avseendet ägnar du särskild uppmärksamhet åt att garantera oskyldighetspresumtionen och säkerställa att offer för diskriminering inte har en oproportionerligt tung bevisbörda (Europarådet, parlamentariska församlingen 2020a).

Dessutom uppmanade resolutionen medlemsländerna att "se till att likabehandlingsorganen har full befogenhet att ta upp frågor om jämlikhet och icke-diskriminering som uppstår från användningen av AI". Den kräver också att regeringar uppmanas att "informera parlamentet innan teknik [för AI- och ADM-system] tas i bruk" och att "myndigheternas användning av sådan teknik ... systematiskt registreras i ett offentligt

register". Det är anmärkningsvärt att många av dessa detaljerade rekommendationer inte återspeglas i den nyligen antagna ramkonventionen om AI. Det pågående arbetet i Steering Committee on Anti-discrimination, Diversity and Inclusion (styrkommittén för antidiskriminering, mångfald och inkludering) och Gender Equality Commission (Jämställdhetskommissionen) efter en rekommendation från Ministerrådet om jämlikhet inom AI kan kanske åtgärda dessa brister (Europarådet, pågående arbete).

Andra instrument som publicerats av institutioner inom Europarådet kan användas för att hantera algoritmisk diskriminering. Till exempel tar den allmänna rekommendationen nr 1 om den digitala dimensionen av våld mot kvinnor, som antogs av expertgruppen för bekämpning av våld mot kvinnor och våld i hemmet (GREVIO) specifikt upp diskriminerande former av våld på nätet, såsom sexuella trakasserier på nätet, stalkning på nätet och med hjälp av teknik samt de digitala dimensionerna av psykologiskt våld (Europarådet 2021a).

Mer övergripande rekommendationer eller instrument som fokuserar på relaterade ämnen är också värda att nämna:

- ▶ Rekommendation 2102 (2017) från Europarådets parlamentariska församling om "Teknologisk konvergens, artificiell intelligens och mänskliga rättigheter" (Europarådet, parlamentariska församlingen 2017).
- ▶ Ministerkommitténs förklaring om risker för de grundläggande rättigheterna till följd av digital spårning och annan övervakningsteknik (Europarådet 2013).
- ▶ Rekommendationen från kommissionären för mänskliga rättigheter om "Uppackning av artificiell intelligens: 10 steg för att skydda de mänskliga rättigheterna" (Europarådet, kommissionären för mänskliga rättigheter 2019).
- ▶ Ministerkommitténs förklaring om algoritmiska processers manipulativa förmåga (Europarådet 2019a).
- ▶ Ministerkommitténs rekommendation CM/Rec(2020)1 till medlemsstaterna om algoritmiska systems inverkan på de mänskliga rättigheterna (Europarådet 2020).
- ▶ Ministerkommitténs rekommendation CM/Rec(2022)13 till medlemsstaterna om den digitala teknikens inverkan på yttrandefriheten (Europarådet 2022b).
- ▶ Den vägledande not som antagits av styrkommittén för medie- och informationssamhället, "Innehållsmoderering. Bästa praxis för effektiva rättsliga och processuella ramverk för självreglerande och samreglerande mekanismer för innehållsmoderering" (Europarådet 2021b).
- ▶ Uppföljningsrekommendationen från kommissionären för mänskliga rättigheter om "Inbyggda mänskliga rättigheter: att framtidssäkra skyddet av mänskliga rättigheter i AI-eran" (Europarådet, kommissionären för mänskliga rättigheter 2023).

Dessa olika instrument ger viktig vägledning när det gäller efterlevnad av mänskliga rättigheter, särskilt icke-diskriminering, vid användning av AI- och ADM-system. Deras icke-bindande karaktär och begränsade verkställbarhet begränsar dock deras effektivitet i praktiken.

Europeiska unionen

På EU-nivå behandlas algoritmisk diskriminering i flera bestämmelser i AI-förordningen (Europeiska unionen 2024c).¹⁰ AI-förordningen är avsedd att komplettera befintliga ramar för antidiskriminering och dataskydd. Det är ett riskbaserat tillvägagångssätt med ett antal förbjudna metoder och högrisksystem som omfattas av särskilda krav. Detta bygger på erkännandet att "[v]id sidan av de många nyttiga användningsområdena för AI kan det också missbrukas och tillhandahålla nya och kraftfulla verktyg för manipulation, utnyttjande och social kontroll ... Sådana användningsområden är särskilt skadliga och kränkande och bör förbjudas eftersom de strider mot unionens värden ... inbegripet rätten till icke-diskriminering" (skäl 28).

Andra system anses utgöra en hög risk eftersom de "kan strida mot rätten till värdighet och icke-diskriminering och värdena jämlikhet och rättvisa" (skäl 31). En tredje kategori av AI-system, som anses vara låg risk, är mycket mer löst reglerade. Med andra ord, "[o]mfattningen av de negativa effekter som AI-systemet har på de grundläggande rättigheter som skyddas av stadgan har särskilt stor betydelse när ett AI-system klassificeras som AI-system med hög risk" (skäl 48). Skäl 27 i rättsakten påminner dock om att de sju principer som fastställts av högnivåexpertgruppen för AI är tillämpliga på alla system oavsett risknivå: dessa är "mänskligt agentskap och mänsklig tillsyn, teknisk robusthet och säkerhet, integritet och dataförvaltning, transparens, mångfald, icke-diskriminering och rättvisa, samhällets och miljöns välbefinnande samt ansvarsskyldighet". I detta sammanhang innebär det att "[m]ångfald, icke-diskriminering och rättvisa innebär att AI-system utvecklas och används på ett sätt som inkluderar olika aktörer och främjar lika tillgång, jämställdhet och kulturell mångfald, samtidigt som diskriminerande effekter och oskäliga biaser, som är förbjudna enligt unionsrätten eller nationell rätt, undviks" (ibid.). Denna princip är tänkt att gälla för alla AI-system som släpps ut på EU-marknaden, oavsett riskkategorisering.

Förbjudna AI-system

Förteckningen över förbjudna system, som är särskilt relevant i samband med diskriminering, ingår i artikel 5 i AI-förordningen. Artikel 5 a förbjuder "subliminal ... eller avsiktligt manipulativa eller vilseledande tekniker [som har] till syfte eller leder till en väsentlig påverkan på en persons eller en grupp av personers beteende genom att avsevärt försämra deras förmåga att fatta ett välgrundat beslut [och] på ett sätt som orsakar eller med rimlig sannolikhet kommer att orsaka betydande skada för [dem]". Artikel 5 b förbjuder AI- och ADM-system som "utnyttjar en sårbarhet hos en fysisk person eller en specifik grupp av personer som härrör från ålder, funktionsnedsättning eller en specifik social eller ekonomisk situation" för att påverka deras beteende och åsamka dem skada.

Artikel 5 c förbjuder sociala poängsättningssystem som utvärderar eller klassificerar personer eller grupper "på grundval av deras sociala beteende eller kända, uttydda eller förutsedda personliga eller personlighetsrelaterade egenskaper" i syfte att

10. Se även motiveringen och den åtföljande AI-pakten som finns på: <https://digital-strategy.ec.europa.eu/en/policies/ai-pact>, hämtad den 10 november 2025.

behandla dem ogynnsamt i sociala sammanhang som saknar koppling till dem eller på ett omotiverat eller oproportionerligt sätt. Artikel 5 d förbjuder användning av ett "AI-system för riskbedömningar av fysiska personer i syfte att bedöma eller förutse risken för att en fysisk person begår ett brott, uteslutande grundat på profileringen av en fysisk person eller på en bedömning av deras personlighetsdrag och egenskaper". Artikel 5 e förbjuder system för ansiktsgenkänning som bygger på skrapning av internet eller CCTV-bilder.

Artikel 5 f förbjuder känsligenkänning på arbetsplatsen och vid utbildningsinstitutioner. Enligt artikel 5 g och h är "användning av system för biometrisk kategorisering som kategoriserar fysiska personer individuellt på grundval av deras biometriska uppgifter för att härleda eller dra slutsatser om en persons ras, politiska åsikter, medlemskap i fackförening, religiösa eller filosofiska övertygelse, sexualliv eller sexuella läggning" förbjuden, liksom "användning av system för biometrisk fjärridentifiering i realtid på allmänt tillgängliga platser för brottsbekämpande ändamål", såvida det inte är absolut nödvändigt.

Dessa förbud tar itu med viktiga risker för algoritmisk diskriminering genom ett generellt förbud. Kritiker har dock pekat på betydande kryphål i de breda undantag som beviljas polis och brottsbekämpande myndigheter, särskilt när det gäller direkt ansiktsgenkänning och biometrisk övervakning i artikel 5 g och h (Europeiska digitala rättigheter 2024). Europeiska kommissionen har utfärdat riktlinjer om förbjudna AI-metoder, som klargör tillämpningsområdet för vissa bestämmelser i AI-förordningen. Riktlinjerna framställer följande:

Även om ett AI-system inte är förbjudet enligt AI-förordningen kan dess användning ändå vara förbjuden eller olaglig enligt annan primär eller sekundär unionsrätt (t.ex. på grund av bristande respekt för de grundläggande rättigheterna i ett visst fall, till exempel bristen på en rättslig grund för behandling av personuppgifter enligt dataskyddslagstiftningen, diskriminering som är förbjuden enligt unionsrätten osv.) (Europeiska kommissionen 2025a, punkt 43).

AI-system med hög risk

I bilaga III till AI-förordningen förtecknas s.k. högrisk-AI som används inom biometri, kritisk infrastruktur (kritisk digital infrastruktur, vägtrafik eller vatten-, gas-, värme- eller elförsörjning), utbildning och yrkesutbildning, rekrytering, arbetsledning och tillgång till egenföretagande, tillgång till och åtnjutande av väsentliga privata tjänster och väsentliga offentliga tjänster och förmåner, brottsbekämpning, migration, asyl och gränskontroll samt rättskipning och demokratiska processer. Dessa högrisk-system omfattas av de specifika rättsliga krav som anges nedan. I artikel 6.3 föreskrivs att ett AI-system som används inom ett område som är förtecknat i bilaga III "[inte ska] betraktas som ett AI-system med hög risk om det inte utgör en betydande risk för skada på fysiska personers hälsa, säkerhet eller grundläggande rättigheter, inbegripet genom att det inte materiellt påverkar resultatet av beslutsfattande". Dessa inbegriper system som "utför en snäv processuell uppgift," eller en rent "förberedande uppgift" och system som är "avsedda att förbättra resultatet av en tidigare slutförd mänsklig verksamhet," eller "att upptäcka beslutsmonster eller avvikelser från tidigare beslutsmonster och [är] inte avsedda att ersätta eller påverka tidigare slutförd mänsklig bedömning, utan ordentlig mänsklig granskning". Denna bestämmelse

kan komma att utesluta ett eventuellt stort antal AI-system som används i högrisk-områden från tillämpningsområdet för artikel 6 och skulle kunna missbrukas för att undvika att uppfylla kraven i AI-förordningen för högrisksystem (se avsnitt 2.5). Sådana system är dock fortfarande föremål för vissa skyldigheter: leverantörer som anser att ett AI-system inte utgör en hög risk måste ändå utarbeta dokumentation för bedömningen innan systemet släpps ut på marknaden eller tas i bruk och måste på begäran tillhandahålla denna dokumentation till nationella behöriga myndigheter. De måste också registrera ett sådant AI-system i den EU-databas som föreskrivs i artikel 71 i AI-förordningen.

AI-modeller för allmänna ändamål

AI-modeller för allmänna ändamål (GPAI), som i artikel 3.63 definieras som AI-modeller "som uppvisar betydande generalitet och på ett kompetent sätt kan utföra ett brett spektrum av specifika distinkta oavsett hur [modellerna] släpps ut på marknaden och som kan integreras i en rad system eller tillämpningar i efterföljande led", är föremål för särskilda bestämmelser i artikel 53 i AI-förordningen. Leverantörerna ska utarbeta och framlägga teknisk dokumentation och bruksanvisningar, följa EU:s upphovsrättslagstiftning och publicera en sammanställning av de uppgifter som används i utbildningssyfte. AI-modeller för allmänna ändamål av en viss kapacitet och omfattning klassificeras som en "systemrisk" enligt artikel 51 i AI-förordningen och omfattas av ytterligare krav enligt artikel 55 i AI-förordningen. Leverantörer av sådana system måste *bland annat* genomföra modellutvärderingar och kontradiktoriska tester samt övervaka och rapportera allvarliga incidenter. När AI-modeller för allmänna ändamål är integrerade i ett AI-system omfattas systemet dessutom av de rättsliga krav som gäller för dess riskkategori (oacceptabel, hög, begränsad eller minimal).

AI-system med begränsade risker

Applikationer som innebär mer begränsade risker omfattas däremot endast av krav på transparens. Till exempel måste distributörer av chattbotar informera användarna om att de interagerar med ett AI-system och distributörer av AI-modeller för allmänna ändamål som producerar text, bilder, videor eller ljud måste markera utdata som artificiellt genererad enligt artikel 50 i AI-förordningen. De återstående AI-applikationerna anses endast medföra minimala risker eller inga risker alls och är oreglerade.

Tidslinje för efterlevnad

AI-förordningen är en förordning, vilket innebär att det inte krävs något införlivande från medlemsstaternas sida: bestämmelserna är direkt tillämpliga. Med detta sagt finns det ett antal milstolpar som är nödvändiga för att lagen ska kunna träda i kraft. Förbud mot användningsområden enligt artikel 5 trädde i kraft i februari 2025 och kommissionen har utfärdat vägledning om dessa förbud (Europeiska kommissionen 2025a). Under 2025 publicerade EU:s AI-byrå en uppförandekod för AI-modeller för allmänna ändamål, inbegripet de som utgör systemrisker (Europeiska kommissionen 2025b), och förväntas ta fram riktlinjer för klassificering av AI-system som hög risk

och en mall för konsekvensbedömningar avseende grundläggande rättigheter för AI-system med hög risk (se avsnittet om konsekvensbedömningar avseende grundläggande rättigheter nedan). I augusti 2025 bör medlemsstaterna ha utsett de nationella behöriga myndigheter (minst en anmälade myndighet och minst en marknadskontrollmyndighet) som har till uppgift att övervaka det nationella genomförandet av AI-förordningen. År 2026 bör EU:s AI-byrå besluta om de tekniska standarder som utarbetats av EU:s standardiseringsorgan, Europeiska standardiseringskommittén (CEN) och Europeiska kommittén för elektroteknisk standardisering (Cenelec), avseende riskhanteringssystem och förebyggande av systematiska fel i AI-system med hög risk (se avsnittet om CEN-Cenelec nedan). Slutligen bör reglerna om högrisksystem träda i kraft i augusti 2026, eller i augusti 2027 för AI-system som används som produkter, eller säkerhetskomponenter därav, som måste genomgå en tredjepartsbedömning av överensstämmelse enligt den harmoniseringslagstiftning som avses i bilaga I till AI-förordningen. För AI-system med hög risk som redan har släppts ut på marknaden eller tagits i bruk senast i augusti 2026 och som är föremål för "betydande ändringar i sitt utförande" eller som är avsedda att användas av offentliga myndigheter, måste leverantörerna och tillhandahållarna vidta nödvändiga åtgärder för att uppfylla kraven och skyldigheterna i AI-förordningen senast i augusti 2030 (artikel 111.2). Alla dessa steg kommer att forma medlemsstaternas förmåga att förebygga och åtgärda algoritmisk diskriminering.

Styrning av data

Vissa bestämmelser i AI-förordningen är direkt relevanta för bekämpningen av algoritmisk diskriminering, särskilt eftersom de innehåller särskilda krav för högrisksystem. I artikel 10 om "data och dataförvaltning" fastställs kvalitetskriterier för utbildning, validering och testning av de högrisksystem som är förtecknade i bilaga III. Det fastställs särskilt i artikel 10.2 f och g att leverantörer av AI- och ADM-system ska genomföra en "undersökning med avseende på eventuella bias som sannolikt kommer att påverka människors hälsa och säkerhet, inverka negativt på grundläggande rättigheter eller leda till diskriminering som är förbjuden enligt unionsrätten, särskilt när utdata påverkar indata för framtida verksamhet" och vidta "lämpliga åtgärder för att upptäcka, förebygga och begränsa eventuella bias som identifierats enligt led f". För dessa ändamål kan de särskilda kategorierna av personuppgifter, som GDPR och AI-förordningen i allmänhet förbjuder att använda, behandlas med stöd av undantaget som anges i artikel 10.5:

I den utsträckning det är absolut nödvändigt för att säkerställa upptäckt och korrigering av bias i samband med ... AI-system med hög risk [och under vissa förutsättningar¹¹], får leverantörerna av sådana system undantagsvis behandla särskilda kategorier av personuppgifter, med förbehåll för lämpliga skyddsåtgärder för fysiska personers grundläggande rättigheter och friheter.

Detta undantag kan vara användbart för att identifiera algoritmisk diskriminering, men det omfattar inte system med låg risk.

Riskhanteringssystem

Enligt artikel 9 i AI-förordningen ska leverantörer av AI-system med hög risk införa ett riskhanteringssystem. Det fastställs att "[e]tt riskhanteringssystem ska inrättas, genomföras, dokumenteras och underhållas för AI-system med hög risk" och att riskhantering måste förstås som "en kontinuerlig iterativ process som planeras och löper under hela livsrytmen". Leverantörerna uppmanas särskilt att identifiera och analysera "de kända och rimligen förutsebara risker som AI-systemet med hög risk kan medföra för hälsa, säkerhet och grundläggande rättigheter när AI-systemet med hög risk används i enlighet med sitt avsedda ändamål". De måste bedöma och utvärdera "de risker som kan uppstå när AI-systemet med hög risk används i enlighet med sitt avsedda ändamål och under förhållanden där det kan förekomma rimligen förutsebar felaktig användning", liksom "andra risker som eventuellt kan uppstå på grundval av en analys av data som samlats in från det system för övervakning efter utsläppande på marknaden som avses i artikel 72". Slutligen måste de vidta "lämpliga och riktade riskhanteringsåtgärder utformade för att hantera de risker som identifierats". Dessa skyldigheter implementeras med hjälp av tekniska standarder som utarbetas av CEN-Cenelec. Den riskhanteringsstandard som kommer att utfärdas av CEN-Cenelec blir därför ett viktigt verktyg för att hantera risker och hot mot de grundläggande rättigheterna, inklusive förebyggande av – och skydd mot – algoritmisk diskriminering. Även om det inte är obligatoriskt kommer efterlevnad av standarder att leda till en presumtion om överensstämmelse för högrisksystem enligt artikel 40 i förordningen, och branschen kommer därför uppmuntras att följa dessa standarder. Det är dock viktigt att nämna att denna presumtion om överensstämmelse uteslutande är begränsad till kraven i AI-förordningen för AI-system med hög risk

11. Villkor gäller: "(a) upptäckt och korrigering av bias kan inte utföras effektivt genom behandling av andra uppgifter, inbegripet syntetiska eller anonymiserade uppgifter, (b) de särskilda kategorierna av personuppgifter omfattas av tekniska begränsningar för vidareutnyttjande av personuppgifter samt säkerhetsåtgärder och integritetsbevarande åtgärder på en nivå som motsvarar den senaste utvecklingen, inbegripet pseudonymisering, (c) de särskilda kategorierna av personuppgifter omfattas av åtgärder för att säkerställa att de personuppgifter som behandlas är säkra, skyddade, omfattas av lämpliga skyddsåtgärder, inbegripet strikta kontroller och dokumentation av åtkomst, för att undvika missbruk och säkerställa att endast personer som är behöriga har tillgång till dessa personuppgifter med lämpliga konfidentialitetsskyldigheter, (d) de särskilda kategorierna av personuppgifter får inte överas, överföras eller på annat sätt göras tillgängliga för andra parter, (e) de särskilda kategorierna av personuppgifter raderas när biasen har korrigerats eller personuppgifternas lagringstid har löpt ut, beroende på vilket som inträffar först, (f) registren över behandling enligt förordningarna (EU) 2016/679 och (EU) 2018/1725 och direktiv (EU) 2016/680 innehåller skälen till varför behandlingen av särskilda kategorier av personuppgifter var absolut nödvändig för att upptäcka och korrigera bias och varför detta mål inte kunde uppnås genom behandling av andra uppgifter."

och inte omfattar EU:s lagstiftning om grundläggande rättigheter eller lagstiftning mot diskriminering.

Övriga krav

Dessa bestämmelser kompletteras av flera stödskyldigheter för leverantörer av AI-system med hög risk, till exempel artikel 11 om teknisk dokumentation, artikel 12 om loggning, artikel 13 om transparens, artikel 14 om mänsklig kontroll, artikel 15 om riktighet, robusthet och cybersäkerhet samt artikel 17 om kvalitetsstyrningssystem. I artikel 15.4 krävs t.ex. att leverantörer av högrisksystem "som fortsätter att lära sig efter det att de har släppts ut på marknaden eller tagits i bruk ... ska utvecklas på ett sådant sätt att risken för att eventuella biasar i utdata påverkar indata för framtida drift (återföring) elimineras eller minskas så mycket som möjligt och... det säkerställs att sådan återföring hanteras på vederbörligt sätt med lämpliga kompensande åtgärder". Dessa skyldigheter kan göra det lättare att tillämpa befintliga bestämmelser om förbud mot diskriminering.

EU-databas

Artikel 49.1 och 49.2 kräver att leverantörer registrerar AI-system i den EU-databas som avses i artikel 71 i lagen. Dessutom krävs det enligt artikel 49.3 att "tillhandahållare som är offentliga myndigheter, unionens institutioner, byråer eller organ eller personer som agerar på deras vägnar, registrera sig, välja systemet och registrera dess användning". Om AI-system används inom brottsbekämpning, migration, asyl och gränskontroll måste registreringen göras i en icke-offentlig del av EU-databasen som endast EU-kommissionen och nationella myndigheter har tillgång till. Enligt artikel 71 ska resten av databasen vara "allmänt tillgänglig på ett användarvänligt sätt". Detta offentliga register kommer att underlätta utredningar av diskriminerande effekter av AI- och ADM-system, inklusive de som används av offentliga förvaltningar, samt potentiella brottsoffers uppgift att fastställa tillräckliga bevis för diskriminering. Även om det är en bra utgångspunkt räcker det dock inte med offentlig registrering i en EU-omfattande databas för att utmana diskriminerande användning av AI- och ADM-system. Dessutom har civilsamhällesorganisationer (CSO:er) uttryckt oro över att allmänheten inte har tillgång till information om brottsbekämpning, migration, asyl och gränskontroll som finns i det offentliga registret (Access Now 2024).

Konsekvensbedömning av grundläggande rättigheter

I artikel 27 föreskrivs att vissa användare – främst användare som är organ som styrs av offentlig rätt, eller privata enheter som tillhandahåller offentliga tjänster, men även användare av AI-system som är avsedda att användas för att bedöma fysiska personers kreditvärdighet eller fastställa deras kreditvärdighet, och AI-system som är avsedda att användas för riskbedömning och prissättning i förhållande till fysiska personer när det gäller liv- och sjukförsäkring – måste genomföra en konsekvensbedömning avseende grundläggande rättigheter (FRIA) innan de använder AI-system med hög risk. Dessa består av en beskrivning av användarens processer, den tidsperiod under vilken och den frekvens med vilken AI-systemet är tänkt att användas, de kategorier av fysiska personer och grupper som sannolikt påverkas av användningen

av AI-systemet i det specifika sammanhanget, de specifika riskerna för skada som sannolikt påverkar dessa grupper, mänskliga tillsynsåtgärder som vidtagits och hur dessa risker hanteras, inklusive genom intern styrning och klagomålsmekanismer. FRIA ska lämnas in till marknadstillsynsmyndigheterna och en sammanfattning ska registreras i EU-databasen (bilaga IV, avsnitt C, punkt 4). Tillgång till FRIA gör det enklare för individer och likabehandlingsorgan att utmana diskriminerande AI- och ADM-system. Dessutom, i enlighet med skäl 96:

tillhandahållare av AI-system med hög risk, särskilt när AI-system används i den offentliga sektorn, involvera relevanta berörda parter, inbegripet företrädare för grupper av personer som sannolikt kommer att påverkas av AI-systemet, oberoende experter och organisationer i det civila samhället i genomförandet av sådana konsekvensbedömningar och utformningen av åtgärder som ska vidtas om riskerna förverkligas

FRIA kompletteras av artikel 60, som fastställer ett ramverk inom vilket leverantörer (med eller utan användare) kan testa AI-system med hög risk under verkliga förhållanden. Även om allvarliga incidenter måste rapporteras till den nationella marknadskontrollmyndigheten (artiklarna 60.7 och 73) och omedelbart åtgärdas genom begränsningsåtgärder, är det oklart i vilken utsträckning relevanta intressenter, t.ex. likabehandlingsorgan, kan få tillgång till testresultat och begränsningsåtgärder, vilket kan påverkas av det nationella sammanhanget. Å ena sidan föreskrivs i artikel 71.4 att "[d]en information [om testning] som registreras i enlighet med artikel 60 ska vara tillgänglig endast för marknadskontrollmyndigheter och kommissionen, såvida inte den potentiella leverantören eller leverantören har gett sitt samtycke till att denna information också görs tillgänglig för allmänheten". I själva verket undantas den testinformation som registrerats enligt artikel 60.4 c från kraven på tillgänglighet och offentlighet. Å andra sidan anges följande i artikel 77.1:

Nationella offentliga myndigheter eller organ som utövar tillsyn över eller kontrollerar efterlevnaden av skyldigheter enligt unionsrätt om skydd av grundläggande rättigheter, inbegripet rätten till icke-diskriminering, i samband med användningen av AI-system med hög risk som avses i bilaga III, ska ha befogenhet att begära och få åtkomst till all dokumentation som skapas eller upprätthålls enligt denna förordning på ett språk och i ett format som är lättillgängligt när åtkomst till sådan dokumentation är nödvändig för att effektivt fullgöra sina mandat inom ramen för deras jurisdiktion. (vår kursivering).

Myndigheter som skyddar grundläggande rättigheter (organ enligt artikel 77)

Faktum är att artikel 77 i AI-förordningen om "Myndigheters befogenheter att skydda grundläggande rättigheter" är särskilt viktig eftersom den fastställer ramen för verkställighet. Som förklarats ovan ger artikel 77.1 tillsynsmyndigheter för grundläggande rättigheter som utsetts av medlemsstaterna befogenhet att vid behov begära och få tillgång till all dokumentation som skapats eller upprätthållits enligt AI-förordningen. Därigenom ges en rad myndigheter som skyddar de grundläggande rättigheterna rätt att få tillgång till dokumentation som upprättats enligt AI-förordningen, till exempel riskhanteringsplaner, konsekvensbedömningar eller annan dokumentation som är nödvändig för att de ska kunna fullgöra sitt uppdrag på ett effektivt sätt. EU-kommissionen har utfärdat en bred tolkning av listan över relevanta myndigheter. Den är inte begränsad till organ som vanligtvis uppfattas som

människorätsstrukturer, till exempel likabehandlingsorgan, nationella institutioner för mänskliga rättigheter och ombudspersoner, utan omfattar även dataskyddsmyndigheter, konsumentskyddsmyndigheter, barnskyddsmyndigheter, arbetsrättsliga myndigheter, tillsynsmyndigheter för media och myndigheter som ansvarar för att säkerställa valintegritet. Myndigheter som beviljas tillträdesrätt enligt artikel 77 måste antingen övervaka eller verkställa relevant unionslagstiftning (eller båda). Kommissionen anser att artikel 77 inte är avsedd att omfatta nationella domstolar eller andra rättsinstanser som agerar i egenskap av domstolar, vilka redan har tillträdesrätt enligt artikel 47 i EU:s stadga om de grundläggande rättigheterna. De nationella offentliga myndigheter eller organ som avses i artikel 77.1 ska tillkännages offentligt och anmälas till Europeiska kommissionen av medlemsstaterna. Med tanke på de nya befogenheter som anförtrotts likabehandlingsorganen (se avsnittet om likabehandlingsorganens roll nedan) kan dessa institutioner komma att spela en viktig roll när det gäller att genomdriva AI-förordningen i förhållande till algoritmisk diskriminering.

I artikel 77.3 föreskrivs dessutom följande:

Om den dokumentation som avses i punkt 1 är otillräcklig för att fastställa huruvida ett åsidosättande av skyldigheter enligt unionsrätt om skydd av grundläggande rättigheter har ägt rum, får den offentliga myndighet eller det offentliga organ som avses i punkt 1 lämna en motiverad begäran till marknadskontrollmyndigheten om organisering av testning av AI-systemet med hög risk genom tekniska medel. Marknadskontrollmyndigheten ska organisera testningen med nära medverkan av den begärande offentliga myndigheten eller det begärande offentliga organet inom rimlig tid efter begäran.

Dessutom föreskrivs i artikel 73.7 i AI-förordningen att utsedda nationella myndigheter som skyddar grundläggande rättigheter ska informeras av en marknadskontrollmyndighet som har mottagit en anmälan om en allvarlig incident, inklusive diskriminering, i samband med ett AI-system med hög risk. Enligt artikel 79.2 ska dessa myndigheter informeras av en marknadskontrollmyndighet, som ska samarbeta med dem till fullo, om den senare identifierar en risk för de grundläggande rättigheterna, inbegripet diskriminering, i samband med ett AI-system som utgör en risk. De berörda operatörerna måste också samarbeta med dem vid behov. Enligt artikel 82.1 ska nationella myndigheter som skyddar grundläggande rättigheter rådfrågas av en marknadskontrollmyndighet som konstaterar att ett AI-system med hög risk som uppfyller kraven utgör en risk för grundläggande rättigheter, inbegripet diskriminering, om den berörda grundläggande rättigheten är relevant för en sådan myndighets eller ett sådant organs uppdrag.

AI-ansvar

År 2022 föreslog EU-kommissionen ett särskilt AI-ansvarsdirektiv som syftade till att undanröja viktiga hinder för tillgång till rättslig prövning i samband med AI. Det föreslogs att ansvarsreglerna skulle harmoniseras i hela EU för civilrättsliga skadeståndsanspråk för skador orsakade av artificiell intelligens som inte grundar sig på avtalsbrott. Det föreslagna direktivet drogs dock tillbaka 2025, vilket gör det ännu mer angeläget att ta itu med frågor om AI-ansvar på nationell och europeisk nivå, inklusive med andra metoder. I utkastet till AI-ansvarsdirektiv erkändes att "AI:s specifika egenskaper ... utgör ett problem för befintliga ansvarsregler" och det

föreslogs att "lätta bevisbördan på ett mycket målinriktat och proportionerligt sätt genom användning av offentliggörande och motbevisbara presumtioner" för att underlätta tillgången till rättslig prövning i fall av skador som orsakats av AI, inklusive algoritmisk diskriminering (EU-kommissionen 2022a). I artikel 3 i det tillbakadragna förslaget anges att en domstol kan ålägga leverantörer eller användare av AI-system med hög risk som misstänks ha orsakat skada att lämna ut bevis när så krävs i rättsliga förfaranden. Om en begäran om att lämna ut eller bevara bevis inte hade följts hade det lett till en motbevisbar presumtion om att svaranden inte hade följt begäran. I motiveringen till artikel 4 i det tillbakadragna förslaget medgavs att "[d]et kan vara svårt för målsägande att fastställa ett orsakssamband mellan bristande efterlevnad och den utdata från ett AI-system som gav upphov till en viss skada", till exempel diskriminering (ibid). Ett åsidosättande av en aktsamhetsplikt enligt AI-förordningen eller andra EU-regler skulle således anses utgöra ett fel och hade utlöst en motbevisbar presumtion om ett orsakssamband mellan skadan och felet. Om t.ex. leverantören av ett AI-system med hög risk inte hade infört ett riskhanteringssystem enligt AI-förordningen, hade en domstol rimligen kunnat anta att det fanns ett orsakssamband mellan detta åsidosättande av aktsamhetsplikten och den skada som systemet orsakade, till exempel diskriminering. Dessa regler hade kunnat underlätta fastställandet av ett prima facie-fall av algoritmisk diskriminering i domstol där svaranden inte hade uppfyllt sina skyldigheter enligt AI-förordningen. Tre villkor skulle dock ha behövt uppfyllas: målsägaren hade behövt visa att det förelåg ett brott mot aktsamhetsplikten, det hade också behövt vara rimligt sannolikt att ett sådant fel eller brott hade påverkat AI-systemets skadliga utdata och målsägaren hade behövt visa att AI-systemets utdata gav upphov till skadan. För enskilda offer av diskriminering hade det redan inneburit betydande hinder att fastställa dessa faktorer. Därför är det avgörande att överväga att anpassa befintliga regler om presumtioner om diskriminering och omvänd bevisbörda i samband med AI- och ADM-system.

Annan EU-lagstiftning och andra riktlinjer

Andra delar av lagstiftningen kan användas för att ta itu med algoritmisk diskriminering på EU-nivå. Exempelvis omfattar artikel 34 i förordningen om digitala tjänster, som gäller riskbedömning av leverantörer av mycket stora onlineplattformar och av mycket stora onlinesökmotorer, risker för diskriminering (Europeiska unionen 2022). Detsamma gäller artikel 35 om begränsningen av dessa risker.¹² Dessutom har EU-kommissionen nyligen inlett ett offentligt samråd om rättsakten om digital rättvisa, som syftar till att stärka skyddet och den digitala rättvisan för konsumenter.

EU har också antagit viktiga policyer om styrning av AI- och ADM-system som kan användas av likabehandlingsorgan, till exempel för att ge policyrekommendationer till beslutsfattare eller öka medvetenheten. Den europeiska förklaringen om digitala rättigheter och principer för det digitala decenniet som antogs 2022

12. Om hur likabehandlingsorgan kan ta itu med AI-driven diskriminering på nätet, inklusive hets mot folkgrupp, se till exempel: Equinet (2018), "Extending the agenda. Equality bodies addressing hate speech", som finns på: <https://equineteurope.org/extending-the-agenda-equality-bodies-addressing-hate-speech>, hämtad den 7 november 2025; Facing Facts Network/CEJI (2022), "Current activities and gaps in hate speech responses: a mapping report for the Facing Facts Network", som finns på: www.facingfacts.eu/hate-speech-report/, hämtad den 7 november 2025.

innehåller till exempel en förklaring om digitala rättigheter (EU-kommissionen 2022b). Deklarationen uttrycker ett åtagande om att "säkerställa att algoritmiska system baseras på lämpliga datauppsättningar för att undvika olaglig diskriminering och möjliggöra mänsklig övervakning av resultat som påverkar människor". EU-kommissionens vitbok från 2020 behandlar också algoritmisk diskriminering.

1.4.2. Instrument för att motverka diskriminering

Även om AI-specifika regler är viktiga nya verktyg för likabehandlingsorgan och andra organisationer som bekämpar diskriminering blir det avgörande att tillämpa lagstiftning mot diskriminering och dataskyddslagstiftningen tillsammans med dessa nya bestämmelser för att ta itu med algoritmisk diskriminering. AI-specifika rättsliga ramverk kompletterar, men ersätter inte på något sätt, jämlikhetslagstiftningen. AI-förordningens krav på AI-system med hög risk innebär i synnerhet en uppsättning tekniska skyddsåtgärder som stöder – men inte kan säkerställa – efterlevnad av lagstiftning mot diskriminering. I detta sammanhang kommer likabehandlingsorganen att spela en viktig roll när det gäller att tillämpa befintliga ramar för jämlikhetslagstiftning på algoritmisk diskriminering. De kan till exempel ge välgrundade råd om den konkreta tillämpningen av reglerna om presumtioner om diskriminering och omvänd bevisbörda vid informationsasymmetrier. De kan också spela en viktig roll när det gäller rättsliga definitioner av algoritmiska skador i form av direkt, indirekt och intersektionell diskriminering, eller ge råd vid frågor om till exempel hur man tillämpar undantag och objektiva motiveringar i detta sammanhang.

Europarådet

På Europarådets nivå finns flera instrument som förbjuder diskriminering och som är tillämpliga på algoritmisk diskriminering.

Viktigast av allt är att artikel 14 i Europeiska konventionen om skydd för de mänskliga rättigheterna, som innehåller ett förbud mot diskriminering, anger att "[å]tnjutandet av de fri- och rättigheter som anges i denna konvention ska säkerställas utan diskriminering på någon grund, till exempel kön, ras, hudfärg, språk, religion, politisk eller annan åsikt, nationellt eller socialt ursprung, tillhörighet till nationell minoritet, förmögenhet, börd eller ställning i övrigt". Denna icke uttömmande lista över förbjudna former av diskriminering kan användas för att hantera de diskriminerande effekterna av AI- och ADM-system när någon av de grundläggande rättigheter som anges i konventionen har överträtts.

Detta kompletteras av artikel 1.1 i protokoll 12 till konventionen, som innehåller ett fristående förbud mot diskriminering: "Åtnjutandet av varje rättighet som anges i lag skall säkerställas utan diskriminering på någon grund, såsom kön, ras, hudfärg, språk, religion, politisk eller annan åsikt, nationellt eller socialt ursprung, tillhörighet till nationell minoritet, förmögenhet, börd eller ställning i övrigt." Dessutom anges i artikel 1.2 att "[i]ngen skall bli diskriminerad av en offentlig myndighet på någon grund såsom dem som nämnts under punkt 1". Detta förbud mot diskriminering gäller – oavsett om någon annan grundläggande rättighet kränks – för offentliga förvaltningar som använder diskriminerande AI- och ADM-system i länder som har ratificerat protokoll 12. Detta ramverk mot diskriminering kompletteras, i fråga om

arbete, av artikel 20 i Europeiska sociala stadgan om rätten till lika möjligheter och likabehandling i frågor som rör anställning och yrkesutövning utan diskriminering på grund av kön.

Andra relevanta instrument inbegriper konventionen om förebyggande och bekämpning av våld mot kvinnor och av våld i hemmet (Istanbulkonventionen) som i artikel 1.1 b syftar till att "bidra till att avskaffa alla former av diskriminering av kvinnor och att främja faktisk jämställdhet mellan kvinnor och män, bland annat genom att stärka kvinnors ställning" och som tolkar våld mot kvinnor som "en kränkning av mänskliga rättigheter och en form av diskriminering av kvinnor". I artikel 4 i ramkonventionen för skydd av nationella minoriteter anges att "all diskriminering på grund av tillhörighet till en nationell minoritet ska vara förbjuden" och detta omfattar "diskriminering, fientlighet eller våld på grund av [en persons] etniska, kulturella, språkliga eller religiösa identitet" (artikel 6). Parallellt med detta skyddar artikel 7.2 i den europeiska stadgan om landsdels- eller minoritetsspråk personer mot diskriminering på grund av "användningen av ett landsdels- eller minoritetsspråk".

Utöver dessa grundläggande instrument kan andra viktiga dokument användas för att bekämpa algoritmisk diskriminering. Till exempel erkänner Europarådets jämställdhetsstrategi 2024–2029 "AI:s särskilda inverkan på jämställdhet och kvinnors rättigheter" och uppmanar "medlemsstaterna [att] ta itu med algoritmisk könsbaserad och intersektionell diskriminering genom människorättsbaserade och mångfacetterade jämställdhets- och icke-diskrimineringsstrategier [och] tillämpa nyutvecklade standarder på området artificiell intelligens och jämställdhet" (Europarådet 2024c). Detta skapar ett mandat för medlemsstaterna att se över sin lagstiftning och åtgärda brister i förhållande till algoritmisk diskriminering.

Ministerkommitténs rekommendation CM/Rec(2019)1 till medlemsstaterna om att förebyggande och bekämpning av sexism erkänner att "internet har gett en ny dimension för uttryck och spridning av sexism, särskilt sexistisk hatpropaganda, till en stor publik, även om sexismens rötter inte ligger i tekniken, utan i ihållande ojämlikhet mellan könen" (Europarådet 2019b). Där står det att "artificiell intelligens utgör särskilda utmaningar när det gäller jämställdhet och könsstereotyper" och att "användningen av algoritmer kan överföra och förstärka befintliga könsstereotyper och därför bidra till att sexism vidmakthålls". Den erkänner AI- och ADM-systemens roll när det gäller att "intensifiera den granskning som kvinnors kroppar, tal och aktivism utsätts för", särskilt på nätet. Rekommendationen uppmanar medlemsstaterna att "integrera ett jämställdhetsperspektiv i all politik, alla program och all forskning som rör artificiell intelligens för att undvika de potentiella riskerna med teknik som vidmakthåller sexism och könsstereotyper och undersöka hur artificiell intelligens kan bidra till att minska klyftorna mellan könen och eliminera sexism". Den rekommenderar att "öka kvinnors och flickors deltagande på informations- och teknikområdet" och kräver att "utformningen av datadrivna instrument och algoritmer ... tar hänsyn till könsbaserad dynamik". Dessutom bör "man förbättra [t]ransparensen kring dessa frågor och öka medvetenheten om den potentiella könsfördelningen i stordata, samt erbjuda lösningar för att förbättra ansvarsskyldigheten". Sammantaget utgör denna rekommendation ett viktigt rättsligt instrument för att bekämpa algoritmisk diskriminering.

Ministerkommitténs rekommendation CM/Rec(2022)17 till medlemsstaterna för skydd av rättigheterna för kvinnor och flickor som migrerar, flyr eller söker asyl innehåller ett avsnitt om AI, ADM och dataskydd. Rekommendationen föreskriver att "all utformning, utveckling och tillämpning av artificiell intelligens och system för automatiserat beslutsfattande inom den offentliga eller privata sektorn eller av tjänsteleverantörer och entreprenörer bör vara icke-diskriminerande, förenlig med principerna om integritet, transparent och ha tydliga styrmekanismer" i samband med gräns- och migrationshantering (Europarådet 2022c). Den uppmuntrar stater att se till att konsekvensbedömningar av mänskliga rättigheter genomförs innan AI- och ADM-system införs på migrationsområdet. Rekommendationen kräver också att civilsamhällesorganisationer (CSO:er) involveras i diskussioner om utveckling och användning av ny teknik som påverkar migrerande, flykting- och asylsökande kvinnor och flickor (ibid.).

Riktlinjerna från Europarådets ministerkommitté om att säkerställa jämlikhet och skydda mot diskriminering och hat under covid-19-pandemin och framtida liknande kriser innehåller flera bestämmelser som tar upp den diskriminerande potential som digitalisering, AI och kontaktsparningsteknik har för "utsatta grupper", även om fokus ligger på digital uteslutning och bristande tillgång (Europarådet 2021c). Rekommendation CM/Rec(2022)16 om bekämpning av hatpropaganda och rekommendation CM/Rec(2024)4 om bekämpning av hatbrott uppmanar medlemsstaterna att "skydda mänskliga rättigheter och grundläggande friheter i den digitala miljön", understryker internetförmedlares roll och ansvar vid spridning av hatpropaganda, betonar vikten av digitala bevis och data vid utredning av hatbrott och erkänner den roll som "extremistrelser som främst verkar genom digitala kanaler och onlinekommunikation" spelar i hatbrott (Europarådet 2022d). Dessa rekommendationer behandlar dock inte fullt ut den roll som algoritmiska rekommendationssystem på sociala medier spelar när det gäller att främja extrem polarisering på nätet och därmed underblåsa diskriminerande stereotyper och fördomar bland användarna.

Europarådets konvention om skydd för barn mot sexuell exploatering och sexuella övergrepp (Lanzarote-konventionen) erkänner "barns såväl som förövares ökade användning av informations- och kommunikationsteknik" och den digitala teknikens roll när det gäller att underlätta övergrepp mot barn (Europarådet 2007). Ministerkommitténs rekommendation CM/Rec(2010)5 om åtgärder för att bekämpa diskriminering på grund av sexuell läggning eller könsidentitet, som antogs 2010, kan användas övergripande för att ta itu med algoritmisk diskriminering, men innehåller inte någon bestämmelse som specifikt tar upp den diskriminerande effekten av AI- och ADM-system, även om allt fler bevis pekar på effekterna av dessa tekniker när det gäller att underblåsa fördomar och uteslutning både online och offline (Europarådet 2010). Detsamma gäller befintliga instrument som förbjuder diskriminering av romer och resandefolk, till exempel ministerkommitténs rekommendation CM/Rec(2024)1 till medlemsstaterna om jämställdhet för romska och resande kvinnor och flickor (Europarådet 2024d).

De allmänna policyrekommendationerna från Europeiska kommissionen mot rasism och intolerans ger också användbara juridiska resurser för att hantera algoritmisk diskriminering, särskilt genom den fältspecifika tillämpningen av AI- och ADM-system:

- Rekommendation nr 6 om bekämpning av spridning av rasistiskt, främlingsfientligt och antisemitiskt material via internet kan bidra till att bekämpa rasistiskt innehåll, hatpropaganda, trakasserier på nätet och extremism som drivs av algoritmiska rekommendationssystem (Europeiska kommissionen mot rasism och intolerans 2000).
- Rekommendation nr 7 om nationell lagstiftning för att bekämpa rasism och rasdiskriminering erbjuder intressanta processuella resurser för att hantera algoritmisk diskriminering på strukturell nivå:

Lagen bör föreskriva att organisationer som föreningar, fackföreningar och andra juridiska personer som, enligt de kriterier som fastställts i den nationella lagstiftningen, har ett legitimt intresse av att bekämpa rasism och rasdiskriminering, har rätt att väcka talan i civilmål, ingripa i förvaltningsärenden eller göra brottsanmälningar, även om det inte hänvisas till ett specifikt offer (Europeiska kommissionen mot rasism och intolerans 2017: 8)

...

[Detta är] viktigt för att kunna ta itu med diskrimineringsfall där det är svårt att identifiera ett offer eller fall som berör ett obestämt antal offer (ibid.): 22).

- Rekommendation nr 8 om bekämpning av rasism i kampen mot terrorism uppmanar medlemsstaterna att

säkerställa att ingen diskriminering följer av lagstiftning och förordningar – eller deras genomförande – särskilt i följande områden:

- Kontroller som utförs av brottsbekämpande tjänstemän i länderna och av gränskontrollpersonal.
- Administrativt frihetsberövande och häktning före rättegång.
- ...
- Rättvis rättegång, brottmålsförfarande.
- Skydd av personuppgifter.
- Skydd av privat- och familjeliv.
- Utvisning, utlämning, deportering och principen om *non-refoulement*.
- Utfärdande av viseringar.
- Uppehålls- och arbetstillstånd samt familjeåterförening.
- Förvärv och återkallande av medborgarskap.

(Europeiska kommissionen mot rasism och intolerans 2004).

Eftersom AI- och ADM-system i allt högre grad används inom dessa områden, särskilt för att underlätta brottsbekämpning, utgör rekommendationen ett användbart ramverk för att hantera deras diskriminerande potential, även om den i sig inte direkt tar upp dessa tekniker.

- Rekommendation nr 10 om bekämpning av rasism och rasdiskriminering i och genom skolundervisningen gäller även när AI- och ADM-system används, t.ex. i urvalsprocesser (Europeiska kommissionen mot rasism och intolerans 2006).

- Rekommendation nr 11 om bekämpning av rasism och rasdiskriminering inom polisväsendet, en sektor där AI- och ADM-system används genom t.ex. teknik för ansiktsgenkänning och ansiktsmatchning, men även brottsförutsägelse- och övervakningssystem, är relevant (Europeiska kommissionen mot rasism och intolerans 2007).
- Rekommendation nr 14 om bekämpning av rasism och rasdiskriminering i arbetslivet ger användbara riktlinjer på ett område där verktyg för cv-granskning och anställningsstöd används i allt större utsträckning (Europeiska kommissionen mot rasism och intolerans 2012).
- Rekommendation nr 15 om bekämpning av hatpropaganda kräver att medlemsländerna "säkerställer att omfattningen av brott [relaterade till hatpropaganda] definieras på ett sätt som tillåter deras tillämpning att hålla jämna steg med den tekniska utvecklingen" (Europeiska kommissionen mot rasism och intolerans 2015).

Detta ramverk kompletteras av andra rekommendationer som fokuserar på specifika former av diskriminering, till exempel rasism, främlingsfientlighet, antiziganism, antisemitism, hat mot muslimer samt diskriminering av migranter som vistas olagligt i landet och av homosexuella, bisexuella, transpersoner och intersexuella personer (Hbti).¹³

Europeiska unionen

På EU-nivå garanteras principerna om icke-diskriminering och jämställdhet i artikel 2 i EU-fördraget om EU:s värden och artikel 3 i EU-fördraget om den inre marknaden, samt i artiklarna 8 och 10 i fördraget om Europeiska unionens funktionssätt (EUF-fördraget) om integrering av jämställdhet och icke-diskriminering, artikel 19 i EUF-fördraget som ger EU mandat att anta lagstiftning mot diskriminering och artikel 157 i EUF-fördraget om lika lön. Artikel 21.1 i EU:s stadga om de grundläggande rättigheterna föreskriver att "[a]ll diskriminering på grund av kön, ras, hudfärg, etniskt eller socialt ursprung, genetiska särdrag, språk, religion eller övertygelse, politisk eller annan åskådning, tillhörighet till nationell minoritet, förmögenhet, börd, funktionshinder, ålder eller sexuell läggning ska vara förbjuden". Artikel 23 i stadgan garanterar att "jämställdhet mellan kvinnor och män måste säkerställas på alla områden, inklusive rekrytering, arbete och lön".

Flera direktiv skapar ett finmaskigt nät av regler mot diskriminering, bland annat:

- Direktiv 2000/43/EG om genomförandet av principen om likabehandling av personer oavsett deras ras eller etniska ursprung.
- Direktiv 2000/78/EG om inrättande av en allmän ram för likabehandling i arbetslivet.
- Direktiv 2004/113/EG om genomförande av principen om likabehandling av kvinnor och män när det gäller tillgång till och tillhandahållande av varor och tjänster.

13. Se alla rekommendationer på: www.coe.int/en/web/european-commission-against-racism-and-intolerance/ecri-standards, hämtat den 9 november 2025.

- ▶ Direktiv 2006/54/EG om genomförandet av principen om lika möjligheter och likabehandling av kvinnor och män i arbetslivet (omarbetning).
- ▶ Direktiv 2010/41/EG om tillämpning av principen om likabehandling av kvinnor och män som är egenföretagare.
- ▶ Direktiv 79/7/EEG om successivt genomförande av principen om likabehandling av kvinnor och män i fråga om social trygghet.

Dessa är tillämpliga på algoritmisk diskriminering förutsatt att diskriminerande AI- och ADM-system används inom deras materiella tillämpningsområde. Denna omfattning uppvisar dock brister, t.ex. när det gäller diskriminering på grund av ålder, sexuell läggning, funktionshinder och religion eller övertygelse vid köp av varor och tjänster.

Europeiska unionens direktiv om likabehandlingsorgan

Två direktiv har nyligen reformerat likabehandlingsorganens uppdrag. Likabehandlingsorganen är viktiga aktörer i kampen mot algoritmisk diskriminering:

- ▶ Direktiv 2024/1499 om om standarder för likabehandlingsorgan på området likabehandling av personer oavsett ras eller etniskt ursprung, likabehandling i fråga om anställning och yrke oavsett personers religion eller övertygelse, funktionshinder, ålder eller sexuella läggning, likabehandling av kvinnor och män i fråga om social trygghet och tillgång till och tillhandahållande av varor och tjänster.
- ▶ Direktiv 2024/1500 om standarder för likabehandlingsorgan på området likabehandling av och lika möjligheter för kvinnor och män i fråga om anställning och yrke.

Senast i juni 2026 måste EU:s medlemsstater anpassa sin nationella lagstiftning till bestämmelserna i de två EU-direktiven om likabehandlingsorgan. När det gäller automatiserade system och AI föreskriver dessa direktiv att "likabehandlingsorgan bör utrustas med lämpliga mänskliga och tekniska resurser. Dessa resurser bör särskilt göra det möjligt för likabehandlingsorganen att dels använda automatiserade system i sitt arbete, dels bedöma sådana system med avseende på deras överensstämmelse med reglerna om icke-diskriminering."

1.4.3. Instrument för dataskydd

Dataskyddslagar är en annan pusselbit i lagstiftningen som likabehandlingsorgan och andra organisationer som bekämpar ojämlikhet i Europa måste tillämpa för att ta itu med algoritmisk diskriminering.

Europarådet

På Europarådsnivå kan lagstiftningen om dataskydd utnyttjas för att hantera algoritmisk diskriminering. Den moderniserade konventionen om skydd för enskilda vid behandling av personuppgifter (konvention 108+), och i synnerhet artikel 6 som begränsar användningen av särskilda kategorier av uppgifter, uppmanar medlemsstaterna att införa "skyddsåtgärder ... som skyddar mot de risker som behandling av känsliga uppgifter kan innebära för den registrerades intressen, rättigheter och

grundläggande friheter, särskilt en risk för diskriminering" (Europarådet 2018). Artikel 11 i ramkonventionen om AI om integritet och skydd av personuppgifter utvidgar dessa begränsningar till "verksamhet under AI-systemens livscykel", särskilt genom att uppmana medlemsstaterna att säkerställa att "enskilda personers rätt till integritet och deras personuppgifter är skyddade, inbegripet genom tillämpliga inhemska och internationella lagar, standarder och ramar" och att "verkningsfulla garantier och skyddsåtgärder har införts för enskilda personer, i enlighet med tillämpliga inhemska och internationella rättsliga skyldigheter". När konvention 108+ träder i kraft kommer den också att tillämpas på områdena för nationell säkerhet och försvar, vilket innebär att dess tillämpningsområde är bredare än den allmänna dataskyddsförordningen (GDPR).

Detta ramverk kompletteras av de riktlinjer om artificiell intelligens och dataskydd som utfärdades 2019 av den rådgivande kommittén för konventionen om skydd för enskilda vid automatisk databehandling av personuppgifter (konvention 108) (Europarådet 2019c). Där betonas det att "[i] alla faser av behandlingen, inbegripet insamling av uppgifter, bör AI-utvecklare, tillverkare och tjänsteleverantörer anta en strategi för inbyggda skydd av mänskliga rättigheter och undvika alla potentiella bias, även oavsiktliga eller dolda, och risken för diskriminering eller annan negativ inverkan på de registrerades mänskliga rättigheter och grundläggande friheter". Det uppmuntrar också till "samarbete ... mellan tillsynsmyndigheter för dataskydd och andra organ som har behörighet i fråga om AI, t.ex. på följande områden: konsumentskydd, konkurrens, arbete mot diskriminering, samt sektoriella regel- och tillsynsmyndigheter och tillsynsmyndigheter för medier".

Dessutom föreskriver riktlinjer om ansiktsgenkänning som antogs 2021 att "[u]tnyttjandet av ansiktsgenkänning endast för att fastställa en persons hudfärg, religiösa övertygelse eller andra övertygelser, kön, ras eller etniska ursprung, ålder, hälsa eller sociala förhållanden bör förbjudas, såvida inte lämpliga skyddsåtgärder föreskrivs i lag för att undvika varje risk för diskriminering" (Europarådet 2021d). Det innehåller också riktlinjer för utvecklare, tillverkare och tjänsteleverantörer när det gäller datauppsättningarnas representativitet. Riktlinjerna, som grundar sina rekommendationer på artikel 5 i konventionen 108+ om uppgifternas riktighet, fastställer att utvecklare, tillverkare och användare "måste undvika felaktig märkning och därigenom testa sina system i tillräcklig omfattning och identifiera och eliminera skillnader i riktighet, särskilt när det gäller demografiska variationer i hudfärg, ålder och kön, och på så sätt undvika oavsiktlig diskriminering". Riktlinjerna uppmanar också de brottsbekämpande myndigheterna att "när de överväger att använda teknik för ansiktsgenkänning i okontrollerade miljöer ... tar hänsyn till hotet mot olika grundläggande rättigheter, inbegripet rätten till dataskydd, integritet, yttrandefrihet, mötesfrihet och rörelsefrihet, eller förbudet mot diskriminering, beroende på den potentiella användningen på olika platser", särskilt genom konsekvensbedömningar av dataskydd.

Europeiska unionen

På EU-nivå kan flera bestämmelser i GDPR spela en viktig roll när det gäller att motverka diskriminering i AI- och ADM-system (Europeiska unionen 2016a). Först och främst hindrar artikel 9 om särskilda kategorier av personuppgifter i princip

leverantörer och användare från att använda sådana uppgifter som variabler (t.ex. etiketter, ingångsvariabler, riskfaktorer) i AI- och ADM-system. Som anges ovan tillåter dock artikel 10.5 i AI-förordningen sådan användning, i samspel med artikel 9.1(g) i GDPR, när syftet är att upptäcka och motverka algoritmisk diskriminering. I skäl 70 i AI-förordningen klargörs detta:

För att skydda andras rätt att slippa diskriminering som kan följa av bias i AI-system bör leverantörerna undantagsvis, i den utsträckning det är absolut nödvändigt för att säkerställa upptäckt och korrigerande av bias i samband med AI-system med hög risk, med förbehåll för lämpliga skyddsåtgärder för fysiska personers grundläggande rättigheter och friheter ... kunna behandla även särskilda kategorier av personuppgifter, av hänsyn till ett viktigt allmänt intresse i den mening som avses i [artikel 9.2(g) i GDPR].

Enligt artikel 22 i GDPR medförs dessutom rätten att inte bli föremål för helt automatiserade beslut. I synnerhet "[d]en registrerade ska ha rätt att inte bli föremål för ett beslut som enbart grundas på automatiserad behandling, inbegripet profilering, vilket har rättsliga följder för honom eller henne eller på liknande sätt i betydande grad påverkar honom eller henne". Denna bestämmelse har använts för att ifrågasätta diskriminerande kreditbedömning i målen mot SCHUFA Holding och Dun & Bradstreet Austria (se avsnitt 1.4.4.). Även om skäl 71 i GDPR ger registrerade rätt att "få en förklaring till det beslut som fattats efter ... bedömning" av ett fullständigt ADM-system, rådde det länge rättslig osäkerhet kring frågan om GDPR gav rätt till en förklaring och den eventuella karaktären och strukturen av en sådan rätt (Wachter. Mittelstadt och Floridi 2017; Malgieri och Comandé 2017; Selbst och Powles 2017). Enligt GDPR har registrerade rätt att få "information som är nödvändig för att säkerställa en rättvis och transparent behandling", artiklarna 13.2 och 14.2, bland annat om "förekomsten av automatiserat beslutsfattande, inbegripet profilering, enligt artikel 22.1 och 22.4 och, arvid det åtminstone i dessa fall ska lämnas meningsfull information om logiken bakom samt betydelsen och de förutsedda följderna av sådan behandling för den registrerade", artiklarna 13.2 f, 14.2 g och 15.1 h. I beslutet för Dun & Bradstreet Austria (se avsnitt 1.4.4.) bekräftade Europeiska unionens domstol förekomsten och omfattningen av rätten till förklaring i artikel 15.1 h i GDPR, som förutsätter tillgång till meningsfull information om hur ett beslut fattades och ger en grund för att överklaga automatiserade beslut enligt artikel 22.3 i GDPR. I artikel 22.1 i GDPR anges att registrerade har rätt att "inte bli föremål för ett beslut som enbart grundas på automatiserad behandling, inbegripet profilering, som har rättsliga följder för [dem] eller på liknande sätt i betydande grad påverkar [dem]". Denna rätt gäller dock inte om automatiserad behandling eller profilering är "nödvändig för att ingå eller fullgöra ett avtal mellan den registrerade och en personuppgiftsansvarige", eller om den registrerade har lämnat "uttryckligt samtycke" enligt artikel 22.2 a c i GDPR. I detta fall krävs det enligt artikel 22.3 i GDPR att den personuppgiftsansvarige vidtar "lämpliga åtgärder för att skydda den registrerades rättigheter, friheter och rättsliga intressen, åtminstone rätten till personlig kontakt med den personuppgiftsansvarige för att kunna uttrycka sin åsikt och bestrida beslutet". Den rätt till förklaring som domstolen nyligen har fastställt kan underlätta tillgången till rättslig prövning, åtminstone i fall av algoritmisk diskriminering baserad på profilering eller processer som till fullo använder ADM.

På det straffrättsliga området är direktiv 2016/680 om skydd för fysiska personer med avseende på behöriga myndigheters behandling av personuppgifter för att förebygga, förhindra, utreda, avslöja eller lagföra brott eller verkställa straffrättsliga påföljder, och det fria flödet av sådana uppgifter, tillämpligt (Europeiska unionen 2016b). Det ger mer begränsade garantier som kanske inte är tillräckliga för att utmana algoritmisk diskriminering i dessa användningsområden.

1.4.4. Rättspraxis från Europeiska domstolen för de mänskliga rättigheterna och Europeiska unionens domstol

Europeiska domstolen för de mänskliga rättigheterna

I sitt beslut i målet *Glukhin v. Russia* 2023 (Glukhin mot Ryssland) bedömde Europeiska domstolen för de mänskliga rättigheterna lagligheten i polisens användning av teknik för ansiktsgenkänning i realtid (Europeiska domstolen för de mänskliga rättigheterna 2023). I det här fallet greps en person sedan han på sina sociala medier lagt ut foton och en video av en fredlig demonstration som han höll i Moskvas tunnelbana, vilket polisen upptäckte under rutinmässig övervakning av internet. För att identifiera honom använde polisen bilder från videoövervakning med CCTV-kameror som filmar offentliga platser kombinerat med biometriska data, i detta fall ansiktsgenkänningsteknik. Domstolen fann att det skett en kränkning av kändes yttrandefrihet enligt artikel 10 och av hans rätt till privatliv enligt artikel 8 i Europeiska domstolen för de mänskliga rättigheterna. Domstolen uttalade särskilt "att användningen av en mycket integritetskränkande teknik för ansiktsgenkänning i samband med att kändes utövar sin rätt till yttrandefrihet enligt konventionen är oförenlig med de ideal och värderingar som präglar ett demokratiskt rättssamhälle och som konventionen är avsedd att upprätthålla och främja". Genom att beakta "svårigheten för den kändes att styrka sina påståenden eftersom den nationella lagstiftningen inte föreskrev någon officiell registrering eller anmälan av användningen av teknik för ansiktsgenkänning", erkänner domstolen också att bristen på insyn i användningen av sådan teknik gör det svårt att lägga fram bevisning och undanröja dessa hinder (ibid., punkt 72).

Europeiska unionens domstol

En begäran om förhandsavgörande från Belgien om dess breda genomförande av EU:s direktiv om passageraruppgifter, inklusive hur automatiseringen av databehandlingssystemet skulle genomföras av nationella myndigheter, ledde till ett beslut i mål C-817/19 *Ligue des droits humains ASBL v. Conseil des ministres* (Europeiska unionens domstol 2022). Fallet fokuserade på spänningen mellan automatiserad databehandling genom maskininlärningsalgoritmer och icke-diskriminering. Domstolen betonade vikten av skyddsåtgärder för automatiserad databehandling och förbjöd användning av självlärande AI-algoritmer för att förhindra diskriminerande resultat. Dessutom får de förutbestämda kriterier som används för att bedöma terroristhot inte baseras på "en persons ras eller etniska ursprung, politiska åsikter, religion eller filosofiska övertygelse, medlemskap i fackförening, hälsa, sexualliv eller sexuella läggning" (ibid.).

I sitt avgörande i målet C-634/21 *SCHUFA Holding* klassificerade Europeiska unionens domstol profilering i form av riskbedömning i förhållande till krediter som en form av ett individuellt automatiserat beslut som omfattas av artikel 22 i GDPR (Europeiska unionens domstol 2023). Inom ramen för den tvist som ligger till grund för detta beslut nekades kändaren ett lån av ett finansinstitut på grundval av den riskbedömning som gjorts av SCHUFA Holding, ett tyskt kreditbedömningsföretag. Frågan var "om fastställandet av [ett sådant] sannolikhetsvärde ... utgör automatiserat individuellt beslutsfattande i den mening som avses i artikel 22.1 i dataskyddsförordningen", med andra ord "ett beslut som enbart grundas på automatiserad behandling, inbegripet profilering, som har rättsliga följder för den registrerade eller på liknande sätt i betydande grad påverkar honom eller henne" (ibid.: punkt 20 och 27). Även om den riskpoäng som SCHUFA tillhandahåller överförs till en tredje part, som i slutändan är ansvarig för att fatta avtalsbeslutet, hävdade domstolen att denna tredje part "i stor utsträckning förlitar sig på detta värde" (ibid.: punkt 48). Domstolen föreslog också att begreppet "beslut" i artikel 22 i GDPR "är tillräckligt brett för att omfatta resultatet av en beräkning av en persons kreditvärdighet i form av ett sannolikhetsvärde avseende den personens förmåga att fullgöra betalningsåtaganden i framtiden" (ibid.: punkt 46). Om "det sannolikhetsvärde som fastställs av ett kreditupplysningsföretag och meddelas en bank spelar en avgörande roll för beviljandet av kredit, måste fastställandet av detta värde därför i sig kvalificeras som ett beslut som i förhållande till en registrerad har 'rättsliga följder för honom eller henne eller på liknande sätt i betydande grad [påverkar] honom eller henne' i den mening som avses i artikel 22.1 i GDPR" (ibid.: punkt 50). Detta är ett viktigt beslut som ger en rättslig grund för att ifrågasätta diskriminering baserad på profilering och andra beslut som till övervägande del baseras på automatiserade riskbedömningar.

Europeiska unionens domstol utvecklade ytterligare de rättsmedel som står till buds för offer för algoritmisk diskriminering i sitt beslut i målet med Dun & Bradstreet Austria (Europeiska unionens domstol 2025). I detta mål hade kändaren nekats ett mobiltelefonavtal av ett österrikiskt telekommunikationsföretag på grundval av en automatiserad kreditbedömning. Den algoritmiska riskpoäng som Dun & Bradstreet, ett företag som specialiserat sig på att tillhandahålla kreditbedömningar, hade angett visade att hon inte hade tillräcklig finansiell kreditvärdighet för att beviljas detta kontrakt, som skulle ha krävt en månatlig betalning på 10 euro (ibid.: punkt 2 och 16). Kändaren vände sig till den österrikiska dataskyddsmyndigheten, som ålade kreditupplysningsföretaget att utlämna "meningsfull information om logiken bakom det automatiserade beslutsfattandet baserat på personuppgifter" till kändaren (ibid.: punkt 17). Under åberopande av affärshemligheter vägrade kreditupplysningsföretaget inledningsvis att förse kändaren med annan information än hennes riskpoäng (ibid.). Man hade endast angett att "vissa sociodemografiska uppgifter om [kändaren] hade 'getts samma vikt' för att fastställa denna riskpoäng, vilket den österrikiska domstolen ansåg vara otillräckligt. I ett försök att verkställa detta beslut utsåg en inhemsk domstol en expert som ansåg att Dun & Bradstreet Austria, för att uppfylla skyldigheten att lämna "meningsfull information", behöver upplysa kändaren om vilka personuppgifter (t.ex. födelsedatum, adress, kön) som behandlats för att generera "faktorer" och "det specifika värde" som tillskrivits kändaren för varje faktor, "den matematiska formel" som poängen baserades på, "de exakta intervall inom vilka samma värde tilldelas olika uppgifter för samma faktor" samt "en lista med poäng

för perioden som omfattar de sex månader som föregick och de sex månader som följde på fastställandet av [sökandens] poäng, som erhållits med hjälp av samma beräkningsregel” för att kontrollera riktigheten (ibid.: punkt 23–25). Enligt den nationella domstolen är det av avgörande betydelse för verkställigheten att käranden har möjlighet att kontrollera att profileringen är korrekt.¹⁴ I den aktuella tvisten, till exempel, ”visade den information som [käranden] fick, inklusive bland annat den poäng som erhöles, att [hon] hade mycket god kreditvärdighet, [medan] den faktiska profileringen ledde till att hon betraktades som icke kreditvärdig” (ibid.: punkt 27). Tillgång till meningsfull information är därför en förutsättning för att käranden ska kunna bestrida diskriminerande profilering eller automatiserade beslut i enlighet med artikel 22.3 i GDPR (se avsnitt 1.4.3.) (ibid.: punkt 55).¹⁵

Europeiska unionens domstol förklarade att ”rätten att erhålla ’meningsfull information om logiken bakom’ automatiserat beslutsfattande, i den mening som avses i [artikel 15.1 h], ska förstås som en rätt till en förklaring av det förfarande och de principer som faktiskt tillämpas för att på automatiserad väg använda den registrerades personuppgifter i syfte att uppnå ett visst resultat, till exempel en kreditprofil”, vilken ”ska tillhandahållas genom relevant information och i en koncis, överskådlig, begriplig och lättillgänglig form” (ibid.: punkt 58–59). Även om domstolen inte angav några relevanta uppgifter, påpekade den att ”enbart meddelandet av en komplex matematisk formel, till exempel en algoritm, eller ... en detaljerad beskrivning av alla steg i det automatiserade beslutsfattandet” inte uppfyller kraven i artikel 15.1 h eftersom ”inget av dessa utgör en tillräckligt koncis och begriplig förklaring” (ibid.: punkt 59). Med andra ord:

”Informationen om den bakomliggande logiken” i automatiserat beslutsfattande, i den mening som avses i artikel 15.1 h i GDPR, måste beskriva det förfarande och de principer som faktiskt tillämpas på ett sådant sätt att den registrerade kan förstå vilka av hans eller hennes personuppgifter som har använts på vilket sätt i det aktuella automatiserade beslutsfattandet, varvid komplexiteten i de åtgärder som vidtas inom ramen för automatiserat beslutsfattande inte kan befria den personuppgiftsansvarige från skyldigheten att lämna en förklaring (ibid.: punkt 61).

För domstolen kan det vara ”tillräckligt transparent och begripligt att informera den registrerade om i vilken utsträckning en ändring av de personuppgifter som beaktats skulle ha lett till ett annat resultat” (ibid.). I förhållande till lagstiftning för att motverka diskriminering är denna rätt till förklaring väsentlig eftersom den kan göra det möjligt för kärende att jämföra sig med verkliga eller hypotetiska jämförelsepersoner, vilket är ett väsentligt inslag i diskrimineringstestet, som dock har blivit mycket otillgänglig i samband med algoritmisk opacitet och epistemologisk fragmentering (se avsnitt 1.4.3.). Slutligen, om företagshemligheter kan äventyras,

14. Den nationella domstolen anförde att ”[o]m artikel 15.1 h i GDPR inte garanterar detta, skulle rätten till tillgång till den registrerades personuppgifter och annan information som föreskrivs i denna artikel bli meningslös och värdelös, särskilt eftersom varje personuppgiftsansvarig i så fall skulle kunna lämna felaktig information” (ibid.: punkt 29).

15. Domstolen uttalade att ”när ett beslut fattas i det specifika sammanhanget enbart på grundval av automatiserad behandling, är huvudsyftet med den registrerades rätt att erhålla den information som föreskrivs i artikel 15.1 h i GDPR att göra det möjligt för honom eller henne att effektivt utöva de rättigheter som tillkommer honom eller henne enligt artikel 22.3 i denna förordning, nämligen rätten att uttrycka sin åsikt om detta beslut och att bestrida det” (ibid.: punkt 55).

”är den personuppgiftsansvarige skyldig att lämna den påstått skyddade informationen till den behöriga tillsynsmyndigheten eller domstolen, som måste göra en avvägning mellan de aktuella rättigheterna och intressena i syfte att fastställa omfattningen av den registrerades rätt till tillgång enligt artikel 15 i GDPR” (ibid.: punkt 76).

I Frankrike har 15 organisationer nyligen lämnat in en överklagan till Conseil d’État avseende ett beslut från CNAF (Caisse nationale des allocations familiales, en avdelning av socialförsäkringen) om att använda en algoritm för riskbedömning för att avgöra vilka förmånstagare som ska utredas för bedrägeri. ADM-systemet visade sig ge upphov till diskriminerande effekter för bland annat kvinnor och människor som lever i fattigdom. De tvistande parterna bad Conseil d’État att hänskjuta frågor till Europeiska unionens domstol, bland annat om algoritmisk diskriminering. Detta mål bör bevakas eftersom det kommer att ge Europeiska unionens domstol möjlighet att pröva algoritmisk diskriminering i ljuset av direkt och indirekt diskriminering.

1.5. Viktiga intressenter

Europeiska AI-byrån

Flera institutioner spelar en roll i kampen mot algoritmisk diskriminering. Nyinrättade Europeiska AI-byrån¹⁶ (som inrättats inom Europeiska kommissionen) stöttar genomförandet av AI-förordningen och kommer att utreda eventuella överträdelser. Enligt artikel 90 kommer Europeiska AI-byrån särskilt att ansvara för att hantera varningar som utlöses av den vetenskapliga panelen av oberoende experter (artikel 68 i AI-förordningen) om AI-modeller för allmänna ändamål (GPAI) som utgör systemrisk. I likhet med en marknadstillsynsmyndighet kommer AI-byrån också ha befogenhet att övervaka och kontrollera efterlevnaden av AI-system som innehåller en GPAI-modell (artikel 75.1 i AI-förordningen). Det kommer också att ske ett samarbete med marknadskontrollmyndigheter för att övervaka, utreda och bedöma efterlevnad för GPAI som används i högriskområden (artikel 75.2 i AI-förordningen). Enligt artikel 75.3 i AI-förordningen ”[o]m en marknadskontrollmyndighet inte kan slutföra sina utredningar av AI-system med hög risk på grund av att den inte fått tillgång till viss information avseende AI-modellen för allmänna ändamål, trots att den har vidtagit alla lämpliga åtgärder för att inhämta den informationen, får den lämna en motiverad begäran till AI-byrån, genom vilken tillgång till denna information ska verkställas”. Mot bakgrund av dessa bestämmelser är det troligt att likabehandlingsorgan som utsetts enligt artikel 77 i AI-förordningen kommer att samarbeta med AI-byrån direkt eller indirekt inom ramen för sitt samarbete med marknadskontrollmyndigheter när de hanterar risker för – eller överträdelser av – rättigheter att motverka diskriminering. Såsom förklaras i avsnitt 1.4.1. krävs det enligt artikel 73.7 i AI-förordningen att de relevanta marknadskontrollmyndigheterna informerar de nationella offentliga myndigheter eller organ som avses i artikel 77.1 i AI-förordningen vid en allvarlig incident som anses åsidosätta skyldigheter ”enligt unionsrätten avsedda att skydda grundläggande rättigheter” enligt artikel 3.49(c) i AI-förordningen. Enligt artikel 79.2 i AI-förordningen ska marknadskontrollmyndigheter ”när risker för grundläggande

16. Se <https://digital-strategy.ec.europa.eu/en/policies/ai-office>, hämtat den 10 november 2025.

rättigheter identifieras” också ”informera och fullt ut samarbeta med de relevanta nationella offentliga myndigheter eller organ som avses i artikel 77.1”. Samråd med myndigheter enligt artikel 77.1 är också obligatoriskt enligt artikel 82.1 i AI-förordningen.

Europeiska styrelsen för artificiell intelligens

Artikel 65 i EU:s AI-förordning upprättar också en europeisk styrelse för artificiell intelligens.¹⁷ AI-styrelsen består av representanter från medlemsstaterna och ger observatörsstatus till Europeiska datatillsynsmannen. AI-styrelsen spelar en väsentlig roll i verkställandet av AI-förordningen, särskilt genom att hjälpa till att ”samordna och säkerställa samarbetet mellan EU:s medlemsstater, i syfte att uppnå ett konsekvent genomförande och en konsekvent tillämpning av AI-akten i hela unionen” (Europeiska kommissionen 2025c). Den utgör en samarbetsplattform för de nationella behöriga myndigheter som ansvarar för tillämpningen av AI-förordningen och underlättar utbytet av teknisk och regulatorisk expertis samt bästa praxis. Den har också till uppgift att inrätta undergrupper för att underlätta samarbetet i verkställighetsärenden, särskilt två permanenta undergrupper som ska utgöra en plattform för samarbete och utbyte mellan marknadstillsynsmyndigheter och anmälande myndigheter, men också för att ge råd om policyfrågor som rör artificiell intelligens.

Rådgivande forum för AI

Genom artikel 67 i AI-förordning inrättas ett rådgivande forum för AI¹⁸ med ansvar för att tillhandahålla teknisk expertis och ge råd och stöd till AI-styrelsens och Europeiska kommissionens arbete. Den rådgivande gruppen består av ”ett balanserat urval av berörda parter, däribland branschen, uppstarts företag, små och medelstora företag, det civila samhället och den akademiska världen” med en balanserad representation av kommersiella och icke-kommersiella intressen samt små och medelstora företag och andra företag.

Vetenskaplig panel med oberoende experter

Genom artikel 68 i AI-förordningen inrättas en vetenskaplig panel med oberoende experter som ska stödja verksamheten för efterlevnadskontroll enligt förordningen. Dess medlemmar består av oberoende experter som väljs ut av kommissionen på grundval av aktuell vetenskaplig eller teknisk expertis inom AI-området som är nödvändig för kommissionens arbete med att

- ▶ stödja tillämpning och efterlevnad av AI-förordningen, särskilt när det gäller AI-modeller för allmänna ändamål (t.ex. genom att varna AI-byrån om eventuella systemrisker, bidra till utvecklingen av verktyg och metoder för att utvärdera AI-modeller för allmänna ändamål och dess kapacitet, ge råd om klassificeringen av AI-modeller för allmänna ändamål med eller utan systemrisk och bidra till utvecklingen av verktyg och mallar)
- ▶ stödja marknadskontrollmyndigheternas arbete på deras begäran

17. Se <https://artificialintelligenceact.eu/article/65>, hämtat den 10 november 2025.

18. Se <https://artificialintelligenceact.eu/article/67/>, hämtat den 10 november 2025.

- ▶ stödja gränsöverskridande marknadskontrollsverksamhet
- ▶ stödja AI-byrån när den utför sina uppgifter inom ramen för unionsförfarandet för skyddsåtgärder.

CEN-CENELEC

CEN-CENELEC gemensamma tekniska kommittén 21 (JTC 21) arbetar med standarder för AI baserat på en standardiseringsbegäran från Europeiska kommissionen.¹⁹ För närvarande arbetar den gemensamma tekniska kommittén 21 med att ta fram en harmoniserad standard som översätter kravet, fastställt i artikel 9 i AI-förordningen, om att leverantörer av AI-system med hög risk ska ha ett riskhanteringssystem på plats.

Likabehandlingsorganens roll

Likabehandlingsorganen har en nyckelroll. Algoritmisk diskriminering sker ofta på sätt som är svåra att upptäcka och bevisa. Detta utgör ett betydande hinder till rättslig prövning för de som drabbas. Personer som utsätts för algoritmisk diskriminering kan ha svårt att förstå hur eller varför de behandlats orättvist, vilket gör det svårt att överklaga besluten. Likabehandlingsorganen spelar därför en avgörande roll när det gäller att stödja brottsoffer och främja ansvarsutkrävande. I detta sammanhang stärker de nyligen antagna direktiven om likabehandlingsorganen (direktiv 2024/1500 och direktiv 2024/1499) likabehandlingsorganens roll och befogenheter i förhållande till deras oberoende, resurser och roll när det gäller att underlätta tillgången till rättslig prövning. De bekräftar också likabehandlingsorganens roll i arbetet med att motverka AI-driven diskriminering:

Det är avgörande att uppmärksamma de möjligheter och risker som användningen av automatiserade system, inbegripet artificiell intelligens, medför. Likabehandlingsorgan bör särskilt ha lämpliga mänskliga och tekniska resurser. Dessa resurser bör särskilt möjliggöra för likabehandlingsorgan att å ena sidan kunna använda automatiserade system för sitt arbete och å andra sidan kunna bedöma systemens efterlevnad av regler om icke-diskriminering. Om likabehandlingsorganet ingår i ett organ som har flera mandat bör de resurser det behöver för att utöva sitt likabehandlingsmandat säkerställas. (Europeiska unionen 2024b: skäl 21; Europeiska unionen 2024a: skäl 22).

I detta sammanhang kan likabehandlingsorgan spela en avgörande roll i att stödja offer, öka medvetenheten, bedriva forskning och förespråka bättre reglering av AI- och ADM-system:

- ▶ *Agera på eget initiativ* och hantera klagomål från offer för algoritmisk diskriminering. Likabehandlingsorgan kan ha utredningsbefogenheter *på eget initiativ*, och de fungerar som en tillgänglig kontaktpunkt för potentiella offer för algoritmisk diskriminering. De kan hantera klagomål från offer, utreda påstådda fall av algoritmisk bias och ge råd och stöd om rättsliga åtgärder.
- ▶ Tillhandahålla rättsligt bistånd och föra ärenden till domstol: detta kan omfatta att avgöra ärenden som rör algoritmisk diskriminering, ge råd till enskilda om

19. Europeiska kommissionens genomförandebeslut av den 22 maj 2023 med standardiseringsbegäran C(2023)3215.

hur de kan ifrågasätta diskriminerande beslut som drivs av AI- eller ADM-system eller, i vissa fall, företräda potentiella offer eller agera för deras räkning i domstol.

- ▶ Öka medvetenheten och informera offer och den bredare allmänheten: likabehandlingsorgan kan spela en nyckelroll när det gäller att utbilda allmänheten om de potentiella skadorna av algoritmisk diskriminering, inbegripet att tillhandahålla information om hur man kan ifrågasätta orättvisa automatiserade beslut och vilka rättigheter individer har i samband med AI-drivna processer och ADM-processer.
- ▶ Samla in uppgifter och bedriva forskning om diskriminering kopplad till AI- och ADM-system: en annan viktig roll för likabehandlingsorganen är att samla in uppgifter om förekomsten av algoritmisk diskriminering. Genom att systematiskt spåra klagomål som rör AI- och ADM-system och bedriva forskning, genomföra datainsamling och övervakning kan likabehandlingsorgan identifiera diskrimineringsmönster och lyfta fram systemfrågor som kan kräva lagstiftningsåtgärder.
- ▶ Ge rekommendationer till beslutsfattare och lagstiftare: sådana uppgifter kan också ge konkreta bevis på behovet av rättsliga reformer (t.ex. när det gäller registrerings- och informationsskyldigheter). Likabehandlingsorganen kan utifrån sin tillsynsverksamhet ge evidensbaserade rekommendationer till beslutsfattare och lagstiftare om hur AI- och ADM-system ska regleras, särskilt med avseende på efterlevnad och relaterade svårigheter.
- ▶ Samarbeta med intressenter för att främja jämlikhet i AI- och ADM-system: likabehandlingsorgan kan samarbeta med ett brett spektrum av intressenter – inbegripet arbetsgivare, tjänsteleverantörer, offentliga institutioner och civilsamhällesorganisationer (CSO:er) – för att uppmuntra antagandet av goda metoder för jämlikhet avseende AI- och ADM-system. Det kan handla om att hjälpa organisationer att utveckla och genomföra jämlikhetsplaner som minimerar risken för algoritmisk bias, genom att ge vägledning om hur man testar AI och ADM-system för diskriminering.



2. Rättsliga brister

I detta avsnitt identifieras brister och behov i de nuvarande rättsliga, processuella och styrningsmässiga ramarna mot bakgrund av diskrimineringsrisker som uppstår till följd av användningen av AI- och ADM-system i offentliga förvaltningar och andra sektorer som omfattas av denna studie. I rapporten sammanfattas också de största hindren, bristerna och behoven som likabehandlingsorganen och andra relevanta intressenter har stött på i sina övervaknings- och tillsynsfunktioner.

2.1. Motivering för införande av AI- och ADM-system

Avsnitt I har visat att användningen av AI- och ADM-teknik blir alltmer utbredd. Flera problem uppstår dock i samband med att de införs. Å ena sidan är de berättelser som främjar införandet av dessa system problematiska och ökar de rättsliga riskerna. AI- och ADM-system framställs ofta som "neutrala" eller åtminstone "mindre partiska" än mänskliga beslutsfattare. Användningen av dem kan motiveras av en vilja att rationalisera beslutsprocesserna. Eftersom det är känt att dessa system i stor utsträckning är partiska kan en sådan berättelse skapa en illusion av att införandet av AI och ADM minskar eller till och med förhindrar diskriminering. Detta ger upphov till risker för att överträdelser av lagstiftning mot diskriminering förbises.

AI och ADM-system lyfts dessutom ofta fram som ett sätt att öka kostnadseffektiviteten i beslutsprocesser, inte minst inom offentliga förvaltningar som är föremål för ökad offentlig granskning i samband med utgiftsnedskärningar och åtstramningspolitik. Detta resonemang leder till att AI och ADM används för specifika ändamål, t.ex. för

att upptäcka bedrägerier eller göra riskbedömningar, vilket gör att de framställs som redskap för kostnadseffektivitet. En sådan berättelse fördunklar åtminstone två viktiga frågor:

- Bör sådan teknik användas för alternativa ändamål, t.ex. för att förbättra tillgången till rättigheter och offentliga tjänster, som studier regelbundet visar är underutnyttjade? (Défenseur des Droits 2024).
- Kan sådana system verkligen göra det möjligt att sänka de offentliga utgifterna med bibehållen kvalitet på de offentliga tjänsterna, när alla kostnader som är förknippade med skyddet av de grundläggande rättigheterna har internaliserats?

Det är t.ex. kostsamt att förhindra att risker för diskriminering förverkligas genom lämpliga test-, övervaknings- och revisionsförfaranden, att utbilda handläggare som hanterar algoritmiska rekommendationer för att säkerställa en effektiv mänsklig tillsyn och att inrätta effektiva informations- och prövningsmekanismer för användare av offentliga tjänster. Sådana kostnader bör rimligen internaliseras fullt ut av de offentliga förvaltningarna när beslut fattas om att införa sådana system. Sådana överväganden bör också få offentliga beslutsfattare att ställa "nollfrågor": när införandet av ett AI- eller ADM-system framstår som alltför riskfyllt eller kostsamt efter det att förebyggande åtgärder och skydd av grundläggande rättigheter har genomförts på ett adekvat sätt, bör detta leda till ett beslut om att inte införa systemet.

2.2. Bristande transparens

Såsom beskrivs i de relevanta avsnitten nedan finns det en allmän brist på systematisk, tydlig och tillgänglig information om utveckling, experimentering och användning av AI- och ADM-system i och utanför offentliga förvaltningar. Användningen av AI- och ADM-system i offentliga förvaltningar kartläggs inte systematiskt, trots pågående planer eller insatser för att skapa databaser över sådan användning (t.ex. det nederländska algoritmregistret).²⁰ Fall av algoritmisk diskriminering är fortfarande mycket få och rör ofta frågor om dataskydd snarare än diskriminering och jämlikhet. De sällsynta fall som finns har ofta tagits upp i domstol med stöd av likabehandlingsorgan och civilsamhällesorganisationer (CSO:er). Bristen på information om när, varför och var AI- och ADM-system används av offentliga förvaltningar – och därmed bristen på rättsliga rapporteringsskyldigheter som hittills varit tillämpliga – gör det dock svårt för dessa aktörer att effektivt och systematiskt övervaka, bedöma och ifrågasätta fall av algoritmisk diskriminering trots sina utredningsbefogenheter. För att effektivt kunna använda dessa utredningsbefogenheter måste de först och främst känna till förekomsten av AI- eller ADM-applikationer. Istället måste likabehandlingsorganen ofta förlita sig på medieundersökningar, specialiserade civilsamhällesorganisationer (CSO:er) eller informella nätverk för att identifiera potentiella fall av algoritmisk diskriminering. Nationella experter rapporterar att förfrågningar om information (t.ex. enligt lagstiftningen om informationsfrihet) regelbundet skickas till offentliga myndigheter i syfte att skapa offentliga register över användningen av AI- och ADM-system i offentliga förvaltningar, men deras legitimitet ifrågasätts ibland, eller så kan de förbli obesvarade när de inte har någon bindande karaktär,

20. Se <https://algoritmes.overheid.nl/en>, hämtat den 10 november 2025.

särskilt när dessa förfrågningar rör känsliga applikationer såsom teknik för digital övervakning. Dessutom kan de sökande ofta inte verifiera den information de får. Ibland används andra kanaler för att få information, till exempel frågor till regeringen som ställs i samarbete med ledamöter i de nationella parlamenten. Effektiviteten i både förfrågningar utifrån principen om informationsfrihet och frågor i parlamentet undergrävs dock av bristen på insyn i användningen av AI- och ADM-applikationer. Vissa likabehandlingsorgan erbjuder positiva exempel i detta avseende: den finska diskrimineringsombudsmannen kan till exempel skicka obligatoriska förfrågningar om information och förtydliganden till offentliga förvaltningar. Sådana befogenheter kan underlätta tillgången till information och insyn när det gäller användningen av AI- och ADM-system och deras potentiella diskriminerande effekter.

Till viss del kommer bristen på information att åtgärdas genom de nya AI-förordningarna. Exempelvis kommer EU:s AI-databas enligt artikel 71 i AI-förordningen att innehålla offentlig information om AI-applikationer med hög risk.²¹ Dessutom krävs enligt artikel 14(a)(b) i ramkonventionen om artificiell intelligens att "relevant information om AI-system som potentiellt kan ha betydande inverkan på de mänskliga rättigheterna och relevant användning av dem dokumenteras, lämnas till organ som är behöriga att få tillgång till den informationen och, när så är lämpligt och tillämpligt, görs tillgänglig för eller meddelas berörda personer", så att dessa personer kan "bestrida det eller de beslut som fattats med användning av systemet eller för vilket eller vilka användningen av systemet i betydande utsträckning legat till grund". Dessa bestämmelser stärks ytterligare genom artiklarna 8 och 9 i konventionen om transparens och tillsyn samt om ansvarsskyldighet och ansvar. Sådana informations- och insyns krav kan möjliggöra offentlig granskning. På samma sätt kräver artikel 16 i konventionen att ett ramverk för risk- och konsekvenshantering antas och upprätthålls (med stöd av den standardiserade konsekvensbedömningen Huderia), och artikel 16.2(f) i konventionen kräver att "risker, faktiska och potentiella konsekvenser samt riskhanteringsmetoden" dokumenteras. Tillgång till sådan information kan hjälpa organ som ansvarar för att skydda grundläggande rättigheter att förebygga och åtgärda algoritmisk diskriminering. I AI-förordningen föreskrivs i sin tur att de myndigheter som utsetts enligt artikel 77 i AI-förordningen ska ha tillgång till teknisk dokumentation och riskbedömningssystem som bygger på AI-förordningen och konsekvensbedömningar avseende grundläggande rättigheter när detta är obligatoriskt enligt artikel 27 i AI-förordningen (begränsat till vissa användare som tillhandahåller offentliga eller väsentliga privata tjänster).

Med detta sagt hade sådana informations- och insyns krav dock kunnat formulerats i vidare termer. För närvarande gäller detta endast AI-system med hög risk, vilket innebär att system som beskrivs som icke hög risk och ADM-system som inte anses falla inom definitionen av AI utelämnas. År 2020 krävde Europarådets parlamentariska församling att regeringar uppmanas att "underrätta parlamentet innan [AI- och ADM-] teknik tas i bruk" och att "myndigheternas användning av sådan teknik ... systematiskt registreras i ett offentligt register" (Europarådet, parlamentariska församlingen

21. AI-system med hög risk, inbegripet biometri, som används inom områdena brottsbekämpning, migration, asyl och gränskontroll, registreras i en icke-offentlig del av databasen som Europeiska kommissionen och specifika nationella marknadstillsynsmyndigheter (t.ex. dataskyddsmyndigheter) har tillgång till.

2020b). Detta återspeglas inte direkt i den nyligen antagna ramkonventionen om AI, som inte kräver att ett offentligt register upprättas. AI-förordningen innehåller i sin tur inga bestämmelser om att nationella parlament eller EU-institutioner ska underrättas när offentliga myndigheter använder AI- eller ADM-system.

2.3. Frågor om tillgång till rättslig prövning

Tillgången till rättslig prövning försvåras på grund av flera informations- och maktasymmetrier. Som förklarats ovan visar ett viktigt resultat av den forskning som genomförts på nationell nivå att bristen på tydlig och konsekvent kartläggning av offentliga förvaltningars användning av AI- och ADM-system äventyrar rätten till icke-diskriminering och dess tillämpning. Att inte veta att ett visst administrativt beslut har fattats med (stöd av) ett AI- eller ADM-system hindrar personer som berörs av sådana beslut från att ens misstänka algoritmisk diskriminering. Även när misstankar uppstår hindrar avsaknaden av generella skyldigheter för offentliga förvaltningar att tillhandahålla information om vilken typ av system som används, vilka beslutsfaktorer som tillämpas och hur de används i beslutsprocessen, brottsoffer från att lägga fram tillräcklig bevisning för att fastställa antaganden om diskriminering vid domstolar.²²

Den europeiska lagstiftningen mot diskriminering föreskriver visserligen att bevisbördan ska delas mellan kärende och svarande för att underlätta för kärende som hävdar att de utsatts för diskriminering. I EU:s antidiskrimineringslagstiftning föreskrivs det att "när personer, som anser sig kränkta genom att principen om likabehandling inte har tillämpats på dem, inför domstol eller annan behörig myndighet lägger fram fakta som ger anledning att anta att det har förekommit direkt eller indirekt diskriminering, skall åligga svaranden att bevisa att det inte föreligger något brott mot principen om likabehandling."²³ På samma sätt överför Europeiska domstolen för de mänskliga rättigheterna bevisbördan för att diskriminering inte har förekommit till svaranden när kärenden lägger fram trovärdiga och samstämmiga fakta som gör gällande att diskriminering kan ha ägt rum: "[B]evis kan följa av samexistensen av tillräckligt starka, tydliga och samstämmiga slutsatser eller av liknande obestridda presumtioner av fakta" (Europeiska domstolen för de mänskliga rättigheterna 2007). I princip underlättar dessa regler kärendens arbete genom att de endast behöver fastställa en presumtion för diskriminering innan bevisbördan kan övergå till svaranden. Det är sedan upp till svaranden att motbevisa presumtionen om diskriminering.

Dessa regler är dock otillräckliga för att hantera makt- och informationsasymmetrier som uppstår när ett beslut fattas med (stöd av) ett AI- eller ADM-system, och kan vara otillräckliga i fall av algoritmisk diskriminering. Å ena sidan rapporterar nationella experter att domstolar ibland sätter en hög tröskel avseende beviskrav för att utlösa en omfördelning av bevisbördan till svaranden. Å andra sidan försvårar algoritmernas otillgänglighet för enskilda kärenden som försöker fastställa en presumtion för algoritmisk diskriminering (se avsnitt 2.2. ovan). Informations- och maktasymmetrier tar sig också uttryck i vad Milano och Prunkl (2024) kallar

22. I viss mån kan denna svårighet mildras genom att rätten till förklaring enligt GDPR aktiveras, även om de processuella hindren och kostnaderna kan bli höga för enskilda offer.

23. Se artikel 8.1 i direktiv 2000/43/EG. Denna bestämmelse finns i samtliga direktiv mot diskriminering: 2000/78/EG, 2004/113/EG och 2006/54/EG.

”epistemologisk fragmentering”, nämligen ”en strukturell egenskap hos algoritmiskt förmedlade miljöer som isolerar individer” och därmed ”hindrar människor och samhällen från att på ett meningsfullt sätt dela och jämföra sina erfarenheter”. Epistemologisk fragmentering ”gör det därför svårare att ... identifiera och konceptualisera framväxande skador i dessa miljöer”, särskilt eftersom den europeiska lagstiftningen mot diskriminering bygger på att offren kan visa att de har utsatts för särbehandling eller en särskild nackdel jämfört med andra personer i en liknande situation (ibid.). Tillsammans gör algoritmisk opacitet och epistemologisk fragmentering det mycket svårt att utmana diskriminerande AI och ADM rättsligt.

Detta innebär att tillgången till rättslig prövning äventyras i praktiken. För att ge ett konkret exempel skapar de ADM-system som används av flera socialförsäkringsmyndigheter för att upptäcka bedrägerier i Europa en kategori av *de facto* ”misstänkta individer” som är mycket diskriminerande men mycket svår för dessa individer att ifrågasätta, vilket får allvarliga konsekvenser för deras liv (indragning av sociala bidrag, återkrav på felaktiga utbetalningar osv.). Även om rätten till förklaring, som nyligen bekräftades av Europeiska unionens domstol i samband med artikel 15.1(h) och artikel 22 i GDPR i fallet *Dun & Bradstreet Austria*, delvis löser problemet (se avsnitt 1.4.4.), kvarstår makt- och informationsasymmetrier, som måste åtgärdas på så sätt att bevisbördan lättas ytterligare för käranden, genom att ge likabehandlingsorgan och relevanta intressenter möjlighet att stödja potentiella offer och genom mer förebyggande arbete.²⁴ Denna fråga underströks redan 2020 av Europarådets parlamentariska församling, som betonade behovet av att ”ägna särskild uppmärksamhet åt att garantera oskuldspresumtionen och se till att diskrimineringsoffer inte får en oproportionerligt tung bevisbörda”, (Europarådet, parlamentariska församlingen 2020b). Man krävde att medlemsländerna ”säkerställer att likabehandlingsorganen har full befogenhet att ta upp frågor om jämlikhet och icke-diskriminering som uppstår till följd av användning av AI”.

I detta perspektiv kan de befogenheter som vissa europeiska likabehandlingsorgan har att främja skydd mot diskriminering på kollektiv och strukturell nivå bidra till att mildra dessa brister. Vissa likabehandlingsorgan kan faktiskt bidra till att förebygga algoritmisk diskriminering genom tillsyns-, utrednings- och sanktionsbefogenheter. Dessutom kan vissa likabehandlingsorgan övervaka användningen av AI- och ADM-system och ifrågasätta diskriminerande resultat i domstol med hänvisning till allmänintresset, utan att något enskilt offer behöver identifieras i förväg (se Europeiska unionens domstol 2008). Även om detta inte är fallet för alla europeiska likabehandlingsorgan, kan sådana befogenheter minska svårigheterna med att behöva förlita sig på att enskilda offer rapporterar diskriminering och dess följder innan utredningar kan inledas. För att vara effektiva måste sådana befogenheter dock backas upp av robusta informationsrättigheter: likabehandlingsorgan måste ha rätt att begära och erhålla relevant och meningsfull information om hur AI- och ADM-system används. Sådana befogenheter bör utvidgas till att omfatta situationer där AI- och ADM-system används för att stödja och underlätta beslutsfattandet, istället för att begränsas till situationer med enbart fullständig automatisering, detta för att

24. Detta är särskilt viktigt eftersom AI-ansvarsdirektivet har dragits tillbaka. Direktivet föreslog att det skulle bli lättare att fastställa presumtioner och omvänd bevisbörda, om än i begränsad omfattning.

återspegla det faktum att hybridbeslutsfattande är den mest utbredda användningsformen. Införlivandet av de nya direktiven om standarder för likabehandlingsorgan senast i juni 2026 är ett bra tillfälle att se över likabehandlingsorganens befogenheter mot bakgrund av dessa nya utmaningar. Dessutom har alternativa rättsprocessvägar, till exempel mekanismer för kollektiva klagomål (t.ex. *actio popularis*, grupptalan), potential att minska problemen med tillgång till rättslig prövning.

2.4. Tillämpning av befintliga och framtida bestämmelser

2.4.1. Institutionellt samarbete

Det tillsynssystem som inrättats enligt AI-förordningen är komplext och lämnar inte mycket utrymme för enskilda klagomål. Även om marknadskontrollmyndigheterna kan ta emot klagomål finns det ingen skyldighet för dem att utreda dem på grundval av detta. I detta perspektiv blir likabehandlingsorganens roll ännu viktigare. Det kontrollsystem som inrättas genom AI-förordningen förutsätter samarbete mellan flera intressenter, nämligen likabehandlingsorgan, andra tillsynsmyndigheter för grundläggande rättigheter som utsetts enligt artikel 77, marknadskontrollmyndigheter och andra relevanta institutioner. Dessutom kommer likabehandlingsorganen att samarbeta med andra institutioner för grundläggande rättigheter som utsetts enligt artikel 77 i AI-förordningen. Enligt några av de institutioner som intervjuades för denna rapport kan det dock finnas betydande utmaningar i detta avseende. Framför allt måste samarbetsmetoder utvecklas och genomföras, och institutionella kommunikationskanaler säkerställas, för att främja ett tillvägagångssätt som bygger på institutionell komplementaritet och intersektionalitet. Vidare måste behöriga myndigheter, inklusive gemensamma kontaktpunkter och marknadstillsynsmyndigheter, ha en adekvat förståelse för de risker för diskriminering som algoritmisk teknik medför och för rättsliga och institutionella metoder för hantering av diskriminering.

2.4.2. Konceptuella luckor

Andra tillämpningsproblem uppstår utifrån de begrepp och förfaranden som ingår i den europeiska lagstiftningen mot diskriminering. Konkret bygger AI, och i synnerhet maskininlärning, på korrelationer, inte orsakssamband. Detta ökar risken för så kallad proxy-diskriminering, som domare kan vara ovilliga att behandla som direkt diskriminering även om det är känt att sådana proxyrelationer korrelerar med skyddade grunder.²⁵ I vissa sammanhang är det t.ex. känt att individers bostadsort har ett starkt samband med ras och etniskt ursprung. På EU-nivå finns inte bosättningsort med på listan över skyddade grunder i lagstiftningen mot diskriminering. Därför kommer sådan algoritmisk proxydiskriminering sannolikt att behandlas som

25. Se Bartoletti I. och Xenidis R för mer information om proxy-diskriminering. (2023), "Study on the impact of artificial intelligence systems, their potential for promoting equality, including gender equality, and the risks they may cause in relation to non-discrimination", som är tillgänglig på: <https://edoc.coe.int/en/artificial-intelligence/11649-study-on-the-impact-of-artificial-intelligence-systems-their-potential-for-promoting-equality-including-gender-equality-and-the-risks-they-may-cause-in-relation-to-non-discrimination.html>, hämtad 10 november 2025, Europarådets utgivning, Strasbourg.

indirekt diskriminering, utom i fall där bostadsorten anses vara "oupplösligt" eller "inneboende" kopplad till en skyddad grund, som ras eller etniskt ursprung.

I detta avseende kan EU:s sekundärrätt som förbjuder diskriminering visa sig vara alltför begränsad eftersom den endast förbjuder diskriminering på en "sluten" lista över skyddade grunder: kön, ras eller etniskt ursprung, funktionshinder, sexuell läggning, religion eller trosövertygelse och ålder. Algoritmiska skador som har samband med andra sociala egenskaper hade endast betraktas som direkt diskriminerande om de anses ha ett "inneboende samband" med någon av dessa skyddade grunder. Sådana skador kan dock behandlas som indirekt diskriminering om de drabbar en grupp som har anknytning till någon av dessa skyddade grunder. Andra algoritmiska skador kan falla utanför det rättsliga skyddet om de inte omfattas av särskilda nationella lagstiftningsramar, som sträcker sig längre än EU-lagstiftningen.

Däremot kan så kallade "öppna" listor, till exempel de skyddade kriterier som fastställs i artikel 21.1 i Europeiska unionens stadga om de grundläggande rättigheterna och artikel 14 i Europeiska konventionen om skydd för de mänskliga rättigheterna, erbjuda mer flexibilitet när det gäller att hantera algoritmisk proxydiskriminering. Detta beror på att de gör det möjligt för domare att konstatera diskriminering "på grund av bland annat" respektive: "kön, ras, hudfärg, etniskt eller socialt ursprung, genetiska särdrag, språk, religion eller övertygelse, politisk eller annan åskådning, tillhörighet till nationell minoritet, förmögenhet, börd, funktionshinder, ålder eller sexuell läggning" och "kön, ras, hudfärg, språk, religion, politisk eller annan åskådning, nationellt eller socialt ursprung, tillhörighet till nationell minoritet, egendom, börd eller ställning i övrigt".

Frågan om hur man kvalificerar algoritmisk proxydiskriminering är viktig. Medan denna kvalificering inte gör någon skillnad enligt konventionens lagstiftning mot diskriminering på grund av att det finns ett enhetligt system med öppna motive-ringar för både direkt och indirekt diskriminering, är problemet annorlunda enligt EU:s lagstiftning mot diskriminering. Där gynnas direkt diskriminering av en striktare rättfärdigandegrund som bygger på stängda undantag. Med tanke på AI- och ADM-systemens bristande transparens, opacitet och komplexitet är det inte säkert att en öppen rättfärdigandegrund som kräver en kontextuell analys är den lämpligaste. Att styrka direkt diskriminering är dock ofta problematiskt i ett algoritmiskt sammanhang. Utöver problemen med rättfärdigande är begreppen direkt och indirekt diskriminering inte helt lämpade för att fånga algoritmisk diskriminering, som blandar egenskaper från båda begreppen (Xenidis 2021).

Den höga graden av differentiering som utförs av AI- och ADM-system ökar också riskerna för intersektionell diskriminering, särskilt när användare kombinerar flera riskkategorier som påverkar deras riskpoäng (t.ex. kön och migrationshistorik i exempel med partiska arbetslöshetsprognoser). Intersektionell diskriminering erkänns dock fortfarande inte fullt ut i EU:s direktiv mot diskriminering (med undantag för det nyligen antagna lönetransparensdirektivet 2023/970) och förbjuds inte på ett enhetligt sätt i nationell lagstiftning, trots den senaste tidens utveckling. Medlemsländerna behöver därför se över sin lagstiftning mot diskriminering och ändra den vid behov för att säkerställa att den omfattar alla fall där direkt eller indirekt diskriminering,

inklusive diskriminering genom association och intersektionell diskriminering, kan orsakas av användningen av AI- och ADM-system.

2.4.3. Begränsningar av insamling och behandling av data

Slutligen begränsar befintliga rättsliga restriktioner för insamling av uppgifter om jämlikhet möjligheterna att effektivt undersöka algoritmisk diskriminering och främja jämlikhet i användningen av AI- och ADM-system. För att effektivt komma till rätta med diskriminering inom AI kan det vara nödvändigt att verka för mer adekvata rutiner för insamling av uppgifter om jämlikhet. Ytterligare frågor om efterlevnad gäller användningen av känsliga kategorier av personuppgifter. Artikel 10.5 i AI-förordningen tillåter undantagsvis behandling av särskilda kategorier av personuppgifter (definierade i artikel 9 i GDPR), till exempel etnicitet eller sexuell läggning, för att upptäcka bias, men denna möjlighet omfattar inte AI-system som faller utanför högriskkategorin. Detta kan innebära betydande begränsningar när det gäller att upptäcka, förebygga och mildra diskriminerande bias i system som anses ha låg risk och som omfattas av tvärgående icke-diskrimineringsskyldigheter. Det är också osäkert om likabehandlingsorgan och andra tillsynsmyndigheter för grundläggande rättigheter kan utnyttja undantagen i artikel 9 i GDPR och artikel 10.5 i AI-förordningen för att bedöma de diskriminerande effekterna av vissa AI-system. Algoritmisk diskriminering kan vara svårare att upptäcka och övervaka när särskilda kategorier av personuppgifter – som överlappar många (men inte alla) av de kategorier som skyddas enligt EU:s och nationell lagstiftning mot diskriminering – inte kan användas för testning, revision och bedömning.

2.5. "Riskbegränsa" AI-system

Den riskbaserade klassificeringen i AI-förordningen har utsatts för kritik. Den förteckning över högrisktillämpningar som finns i bilaga III till lagen har beskrivits som alltför rigid med tanke på den snabba tekniska utvecklingen. AI-förordningen ger inte heller någon tydlig vägledning för att avgöra om en tillämpning som faller inom de sektorer som anges i bilaga III har hög risk eller inte. Detta är särskilt problematiskt mot bakgrund av artikel 6.3 i AI-förordningen, som föreskriver ett antal undantag. Leverantörer kan t.ex. förklara att system som normalt omfattas av listan över högrisksektorer i själva verket inte utgör någon betydande risk för skada på fysiska personers hälsa, säkerhet eller grundläggande rättigheter. AI-system anses inte utgöra en hög risk när sådana system används för att "utföra en snäv processuell uppgift" eller en "förberedande uppgift för en bedömning",²⁶ för att "förbättra resultatet av en tidigare genomförd mänsklig verksamhet" eller för att "upptäcka beslutsmonster eller avvikelser från tidigare beslutsmonster ... [där sådana system inte är avsedda att] ersätta eller påverka tidigare slutförd mänsklig bedömning, utan ordentlig mänsklig granskning". Leverantörer som anser att deras system inte utgör en hög risk behöver bara "upprätta dokumentation för sin bedömning innan systemet släpps ut på marknaden eller tas i bruk" (artikel 6.4) och registrera systemet i enlighet med artikel 49.2.

26. Enligt artikel 6.3 d anses dock ett AI-system som används för att profilera fysiska personer på de områden som anges i bilaga III alltid utgöra en hög risk.

Det finns ingen systematisk kontroll vid denna punkt eftersom nationella behöriga myndigheter kan, men inte har någon systematisk skyldighet att, begära att bedömningarna utförs. Detta skapar således ett potentiellt kryphål som skulle kunna användas i stor utsträckning för att "avlägsna risken" för AI-system för att undvika att uppfylla de krav och skyddsåtgärder som gäller för högrisksystem enligt AI-förordningen.²⁷ Här kan tillvägagångssättet i ramkonventionen om AI, som är rättighetsbaserat i motsats till riskbaserat, visa sig vara särskilt användbart för att utvidga skyddet av grundläggande rättigheter till AI- och ADM-tillämpningar med "begränsad risk". En annan potentiell lucka som kan utnyttjas i båda regelverken gäller definitionen av AI. Eftersom båda texterna gäller AI i motsats till t.ex. enkla regelbaserade system, kan vissa ADM-system uteslutas om de anses falla utanför omfattningen för AI-definitionen (Europeiska kommissionen 2025d). Detta är ytterligare ett kryphål som leverantörer av enklare ADM-system kan använda för att kringgå skyldigheterna enligt AI-förordningen.

2.6. Tekniska standarder, likabehandlingsorgan och frågan om harmonisering

De harmoniserade standarder för AI som utfärdas av CEN-Cenelec blir ett viktigt verktyg för att stödja efterlevnaden av grundläggande rättigheter, inklusive förebyggande av och skydd mot algoritmisk diskriminering. Den viktigaste frågan är dock vilka risker för diskriminering som standarderna kommer kräva att leverantörerna identifierar, utvärderar och åtgärdar. I synnerhet, vilka former av diskriminering måste leverantörerna ta hänsyn till och hur måste de hantera dem för att anses följa AI-förordningen? Här uppstår flera problem. Den första handlar om den makt som anförtros de privata organisationer som deltar i samregleringsprocessen. De flesta av de organisationer som deltar i JTC 21 inom CEN-Cenelec, som ansvarar för att utarbeta dessa standarder, är branschaktörer som har resurser att finansiera ett sådant deltagande. Civilsamhällesorganisationer (CSO:er) och organisationer med ett legitimt intresse kan delta i viss utsträckning (Equinet är t.ex. för närvarande involverat i standardiseringsprocessen), men har ofta begränsade ekonomiska medel för att göra det. Detta upplägg innebär också att expertisen på fältet kan vara ojämnt fördelad, med en tillräcklig nivå av teknisk expertis och mer begränsad juridisk expertis, särskilt beträffande mänskliga rättigheter. Utkasten till standarder är inte offentliga, vilket gör det svårt att övervaka och bedöma effektiviteten och bristerna i de verkställighetsstrategier som för närvarande diskuteras. Detta förhindrar försök att påverka dessa diskussioner med syfte att ta itu med befintliga begränsningar.

Det andra problemet är huruvida detta synsätt är förenligt med och kompletterar EU:s lagstiftning om grundläggande rättigheter, inklusive lagstiftningen mot diskriminering. Å ena sidan är det fortfarande oklart vilken nivå av riskreducering som standarderna kommer att föreskriva i efterlevnadssyfte. Detta är en nyckelfråga eftersom det i artikel 9.5 i AI-förordningen föreskrivs att "[r]iskhanteringsåtgärderna ... ska vara sådana att relevanta kvarvarande risker förknippade med varje fara samt den totala kvarvarande risken i AI-systemen med hög risk bedöms vara acceptabla". När det gäller bias och diskriminering kommer acceptabla kvarvarande risker enligt

27. Dessa krav och skyddsåtgärder anges i avsnittet om AI-system med hög risk.

AI-förordningen inte att garantera efterlevnad av EU:s lagstiftning om grundläggande rättigheter, som har ett annat synsätt med fokus på faktiska skador. Artikel 52.1 i EU-stadgan om de grundläggande rättigheterna föreskriver att "[v]arje begränsning i utövandet av de rättigheter och friheter som erkänns i denna stadga ska vara föreskriven i lag och förenlig med det väsentliga innehållet i dessa rättigheter och friheter" och att "[b]egränsningar får, med beaktande av proportionalitetsprincipen, endast göras om de är nödvändiga och faktiskt svarar mot mål av allmänt samhällsintresse som erkänns av unionen eller behovet av skydd för andra människors rättigheter och friheter". Omfattningen av tillåtna ingrepp – som begränsas av de grundläggande rättigheterna och är beroende av nödvändighet och allmänintresse – förefaller stå i strid med den metod med kvarvarande risk som antagits i AI-förordningen. Åtminstone är förhållandet mellan dessa begrepp tvetydigt och kommer att kräva förtydligande, eventuellt av EU-domstolen.

Å andra sidan står EU:s standardiseringsstrategi, i kombination med den (åtminstone delvis) maximala harmoniseringsram som fastställts i EU:s nya AI-förordning, i motsättning till metoden med en minimiharmonisering som fastställts i EU:s direktiv mot diskriminering. EU:s sekundärrätt förbjuder sex diskrimineringsgrunder (ras och etniskt ursprung, kön, funktionsnedsättning, sexuell läggning, religion eller övertygelse samt ålder), men tillåter också medlemsstaterna att gå längre än dessa minimikrav och förbjuda diskriminering i större utsträckning genom så kallade "mer gynnsamma bestämmelser", så länge de är förenliga med EU-fördragen. Det är dock oklart vilka former av bias som leverantörer av högrisksystem kommer att behöva testa för. Om till exempel testning och begränsning av bias endast krävs i förhållande till de sex grunder som skyddas på EU-nivå, och leverantörer som uppfyller dessa standarder sedan antas uppfylla AI-förordningen, kommer detta att göra det möjligt för dem att sälja sina produkter på hela den inre marknaden. Även om användare fortfarande kan hållas ansvariga för diskriminering enligt EU-lagar och nationell lagstiftning mot diskriminering som kräver ett bredare skydd, uppstår frågor i åtminstone två avseenden.

För det första är mer gynnsamma nationella bestämmelser om förbud mot diskriminering uttryckligen tillåtna enligt EU:s lagstiftning mot diskriminering och bör inte betraktas som hinder för den fria rörligheten. Ett sådant tillvägagångssätt skulle också strida mot artikel 21.1 i EU-stadgan om de grundläggande rättigheterna, som förbjuder diskriminering "på grund av bland annat kön, ras, hudfärg, etniskt eller socialt ursprung, genetiska särdrag, språk, religion eller övertygelse, politisk eller annan åskådning, tillhörighet till nationell minoritet, förmögenhet, börd, funktionshinder, ålder eller sexuell läggning". Ett annat scenario skulle kunna vara att inkludera den bredare definitionen av diskriminering som finns i stadgan. Spänningarna med en mer skyddande nationell lagstiftning, som uttryckligen tillåts enligt EU:s lagstiftning mot diskriminering, hade dock kvarstått. Om standarderna samtidigt inte nämner de förbjudna diskrimineringsgrunder som leverantörer av AI-system med hög risk måste ta hänsyn till när de testar för bias eller bedömer systemets inverkan på de grundläggande rättigheterna, finns det en betydande risk för att sådana förfaranden i slutändan kommer att vara i stort sett ineffektiva.

Den andra frågan gäller vilken inverkan sådana efterlevnadsmekanismer har på ändringen av bevisbördan i diskrimineringsmål. Här är det viktigt att komma ihåg

att det antagande om efterlevnad som föreskrivs i artikel 40 i AI-förordningen endast gäller lagens krav för högrisksystem (se avsnittet om riskhanteringssystem), och inte att motverka diskriminering eller annan lagstiftning om grundläggande rättigheter. I detta avseende erkänns det i artikel 82.1 i AI-förordningen att "ett AI-system med hög risk uppfyller kraven i denna förordning men ändå utgör en risk för personers hälsa och säkerhet, för grundläggande rättigheter".

Det måste därför understrykas att AI-förordningens (partiella) största möjliga harmonisering inte ska tolkas så att den åsidosätter eller minskar befintliga skyldigheter för motverkande av diskriminering enligt EU:s primär- och sekundärrätt.

Vad innebär detta för likabehandlingsorgan och relevanta myndigheter som skyddar grundläggande rättigheter, t.ex. de som anges i artikel 77 i AI-förordningen? För det första kan det vara svårt att bedöma om riskhanteringssystem och konsekvensbedömningar avseende grundläggande rättigheter är förenliga med lagstiftningen mot diskriminering om de inte innehåller information om de diskrimineringsgrunder som har testats. Den tvingande frågan är om likabehandlingsorgan och relevanta intressenter som upprätthåller grundläggande rättigheter kommer att ha tillräckligt exakt information för att bedöma hur diskriminering har förhindrats och/eller risker för detta har minskats i förhållande till rättsligt skyddade grupper i samband med AI-system med hög risk, och hur dessa grupper har definierats under riskbedömningsförfarandena. För det andra skapar den rättsliga dissonans som lyfts fram ovan en klyfta mellan leverantörernas (och i förekommande fall tillhandahållarnas) skyldigheter att förebygga och hantera risker för diskriminerande bias i samband med högrisksystem, och mandatet för likabehandlingsorgan och andra relevanta myndigheter. De flesta likabehandlingsorgan har mandat att ta itu med diskriminering på alla grunder som skyddas av nationell lagstiftning, vilket ofta sträcker sig längre än de grunder som skyddas av EU:s sekundära lagstiftning mot diskriminering. Beroende på vilket tillvägagångssätt som föredras i de nya standarderna eller vid tillämpningen av dem (t.ex. minsta gemensamma nämnare eller ospecificerade typer av bias) måste likabehandlingsorganen vara särskilt vaksamma när det gäller diskrimineringsgrunder som endast är förbjudna enligt nationell lagstiftning. Det är inte säkert att leverantörerna har testat för risker för algoritmisk diskriminering i förhållande till dessa nationellt skyddade grunder.

Dessutom innebär standardisering flera utmaningar för likabehandlingsorganen. Det kan vara viktigt för likabehandlingsorgan att övervaka och påverka hur nya harmoniserade AI-standarder kommer att definieras av CEN-Cenelec, eftersom efterlevnad av dessa standarder kommer att leda till ett antagande om efterlevnad av de väsentliga kraven i AI-förordningen för leverantörer av AI-system med hög risk. Flera svårigheter uppstår dock. Även om Equinet deltar i vissa undergrupper till den gemensamma tekniska kommittén JTC 21, som ansvarar för att utarbeta dessa AI-standarder, är det fortfarande oklart i vilken utsträckning de överenskomna standarderna offentliggörs och hur tillgängliga de är. I fallet *Public.Resource.Org och Right to Know v. Commission and Others* beslutade EU-domstolen att standarderna måste göras allmänt tillgängliga för EU-medborgarna (Europeiska unionens domstol 2024). Eftersom tvisten gällde tillgång till standarder för leksakssäkerhet kvarstår frågan om detta krav på öppenhet kommer att gälla på samma sätt för AI-standarder och om de kommer att göras tillgängliga kostnadsfritt.

För det andra kan det krävas teknisk expertis för att förstå och kritiskt utvärdera om AI-standarder i tillräcklig utsträckning säkerställer efterlevnad av grundläggande rättigheter och lagstiftning mot diskriminering. Enligt en rapport från Equinet kommer standarderna "att avgöra vilken information och detaljnivå som ska ingå i dokumentationen" som tas fram enligt de grundläggande kraven för AI-system med hög risk enligt AI-förordningen och därmed "den typ av riskhantering och eftermarknadskontroll som leverantörerna ska utföra för att förutse, identifiera och minska riskerna för de grundläggande rättigheterna" (Mittelstadt 2025).

Likabehandlingsorganen bör därför vara särskilt uppmärksamma på de tekniska val som görs i de nya AI-standarderna:

- ▶ Vilka rättvisemått används för att mäta prestandan? Vilka tröskelvärden används för att kvantifiera risker för diskriminerande bias? Vilka åtgärder för att minska bias och vilka rättvisemetoder används? Stämmer de överens med europeisk lagstiftning mot diskriminering och rättspraxis?
- ▶ Hur definieras acceptabla nivåer av "kvarvarande risk" för diskriminering?
- ▶ Hur definieras systemens avsedda ändamål, förväntade noggrannhetsnivå och förutsebara missbruk?
- ▶ Hur mäts och minskas risker för diskriminering och dess effekter med avseende på grupper som skyddas av europeisk lagstiftning mot diskriminering?
- ▶ Hur definieras särskilda grupper i den tekniska dokumentationen, riskhanteringen och konsekvensbedömningarna?
- ▶ Har intersektionell diskriminering beaktats på ett adekvat sätt?
- ▶ Omfattar de tekniska gruppdefinitionerna hela den befolkning som skyddas av de förbjudna diskrimineringsgrunderna?

Det är viktigt att vara uppmärksam på dessa frågor för att förhindra att etiska och grundläggande rättigheter tvättas bort.

Detta är en utmaning för likabehandlingsorganen som kan behöva söka ny expertis, kunskap och resurser för att ta itu med dessa frågor. Att upprätta robusta samarbetsmallar med relevanta myndigheter (marknadskontrollmyndigheter, gemensamma kontaktpunkter, AI-byrån osv.) och organ med kompletterande sakkunskap (data-skyddsmyndigheter, ombudspersoner, konsumentverk osv.) samt relevanta tredje parter (forskare, civilsamhällesorganisationer [CSO:er] osv.) kan hjälpa likabehandlingsorganen att ta sig an utmaningen med sakkunskap och resurser.

Sammanfattning – Användning av AI i belgisk offentlig förvaltning och policyrekommenda- tioner om algoritmisk diskriminering



Sammanfattningen är baserad på rapporten om rådande förhållanden och analys av brister, där den belgiska delen författades av Ine van Zeeland.²⁸

Användning av AI inom offentlig förvaltning

Den belgiska offentliga förvaltningen använder flera applikationer för artificiell intelligens (AI) och har flera pågående AI-projekt, till exempel inom polis-, finans-, rekryterings- och utbildningssektorerna. Eftersom det inte finns något heltäckande offentligt register över dessa system är det inte möjligt att göra en fullständig och sammanhängande kartläggning av användningen av AI och automatiserat beslutsfattande (ADM) inom offentliga tjänster i Belgien. Federal Public Service Policy and Support (FPS BoSA) är värd för ett AI-observatorium som har inlett arbetet med att samla in information om användningen av AI- eller ADM-system i offentliga tjänster i Belgien. Liknande insatser för att samla in information genomförs eller planeras av andra organ, till exempel FARI Institute for Brussels och FPS sociala trygghetssystem.

Lagstiftning och relevant rättspraxis

EU:s AI-förordning utgör tillsammans med europeisk och belgisk lagstiftning mot diskriminering och dataskyddslagstiftning ett rättsligt ramverk för att hantera algoritmisk diskriminering. Belgiens rättspraxis om AI, ADM och diskriminering är relativt ny. I rapporten om rådande förhållanden och analys av brister identifierades några fall som var relevanta för ämnet, till exempel domstolens beslut i C-817/19 *Ligue des droits humains ASBL v. Conseil des ministres*,²⁹ som tog ställning till spänningen

28. Detta är en konfidentiell rapport som granskar hur AI- och ADM-system används inom den offentliga sektorn i Belgien, utvärderar AI-relaterade rättsliga och politiska ramverk ur ett jämställdhets- och icke-diskrimineringsperspektiv och ger rekommendationer till myndigheter och beslutsfattare.

29. C-817/19 *Ligue des droits humains ASBL v. Conseil des ministres* EU:C:2022:491.

mellan automatiserad databehandling genom maskininlärningsalgoritmer och icke-diskriminering.

När det gäller dataskydd har tvistemålsavdelningen GBA-APD avgjort två mål som rör ADM-system med potentiell relevans för motverkande av diskriminering; det ena målet gällde kreditbedömning av ett bildelningsföretag³⁰ och det andra gällde utvecklingen av en algoritmisk modell baserad på betalkortsdata.³¹ I båda fallen avsåg dock besluten i tvistemålsavdelningen GDPR-krav som inte hade någon direkt koppling till diskriminering. Överlag kan det finnas oupptäckta rättsfall där AI- och ADM-system ger upphov till ojämlik behandling. År 2022 avgjorde förstainstansrätten i Bryssel ett mål om indirekt diskriminering genom ADM via skannade bilar som används av parkeringsbolag.³²

Utöver EU:s regelverk ges ytterligare möjligheter att förebygga och hantera algoritmisk diskriminering i Europarådets ramkonvention om artificiell intelligens och mänskliga rättigheter, demokrati och rättsstatsprincipen, som öppnades för undertecknande i september 2024, där artikel 10 om jämlikhet och icke-diskriminering förbjuder diskriminering och kräver att stater "upprätthålla åtgärder som syftar till att övervinna ojämlikheter för att uppnå skäliga, riktiga och rättvisa resultat".

Brister i rättsliga och politiska ramverk

Bristande transparens

En kritisk fråga i Belgien är bristen på transparens och rapportering om användningen av AI- och ADM-system, eftersom det för närvarande inte finns någon formell skyldighet för organisationer inom den offentliga sektorn att offentliggöra eller registrera sin användning av AI- och ADM-system. Avsaknaden av en heltäckande och offentligt tillgänglig databas innebär risker för både tillsyn och ansvarsskyldighet, och skapar dessutom utmaningar för likabehandlingsorgan att övervaka risker för diskriminering eller att stödja individer som riskerar att diskrimineras av sådana system. Forskning på området visar genomgående att fall av algoritmisk diskriminering är underrapporterade och underutredda.

Brist på medvetenhet och ett harmoniserat tillvägagångssätt för grundläggande rättigheter

Trots olika utbildningar och etiska riktlinjer är den systematiska kunskapen om risker i samband med diskriminering begränsad inom många offentliga tjänster. Överlag saknar den offentliga förvaltningen ett standardiserat tillvägagångssätt för att bedöma risker som rör grundläggande rättigheter och diskriminering vid användning av AI i offentliga tjänster. Det saknas ett harmoniserat tillvägagångssätt för att inkludera

30. Décision 168/2023 du 19 décembre 2023, Chambre Contentieuse GBA-APD.

31. Décision quant au fond 46/2024 du 15 mars 2024, Chambre Contentieuse GBA-APD.

32. Mål 21/6495/C, tillgängligt på: www.unia.be/fr/legislation-et-jurisprudence/jurisprudence/tribunal-de-premiere-instance-de-bruxelles-le-2-mai-20221, hämtat den 10 november 2025. Kärande i målet var en person med funktionsnedsättning och Collectif accessibilité Wallonie Bruxelles (CAWaB), med stöd av det interfederala likabehandlingsorganet Unia.

perspektiv från berörda grupper eller individer och experter på grundläggande rättigheter i utformning, utveckling och tillhandahållande av AI- och ADM-system.

Fragmenterat landskap för tillsyn

Belgien har ett institutionellt komplext tillsynslandskap, där mandaten är fördelade på federal, regional och lokal nivå. Detta ger upphov till en utmaning för harmoniserad tillsyn och rättsliga tolkningar och en risk för inkonsekvent tillämpning av AI-förordningen inom offentliga tjänster i Belgien.

Resurs- och kompetensbrister

Likabehandlingsorgan och tillsynsmyndigheter i Belgien står inför betydande utmaningar när det gäller att ta itu med AI- och ADM-driven diskriminering på grund av begränsade offentliga finansiella, mänskliga och tekniska resurser.

Viktiga rekommendationer

- ▶ **Upprätta ett offentligt register över AI- och ADM-system:** Införa en lagstadgad skyldighet att rapportera användningen av AI- och ADM-system i offentlig sektor till ett nationellt register eller en nationell databas med en viss nivå av offentlig tillgång, utöver det EU-omfattande registret, för att säkerställa transparens och effektiv tillsyn.
- ▶ **Inrätta avancerade samordningsmekanismer mellan nationella behöriga myndigheter på olika nivåer:** Underlätta och ta fram formella samarbetsavtal för att garantera harmoniserad tillsyn och beslutsfattande av respektive myndigheter och skapa riktlinjer för ömsesidigt engagemang för att säkerställa att myndigheter enligt artikel 77 i AI-förordningen och nationella strukturer för mänskliga rättigheter proaktivt inkluderas i utredningar.
- ▶ **Bilda ett nätverk för att dela med sig av lovande tillvägagångssätt och öka synergier:** Underlätta och stödja offentliga myndigheter i inrättandet av ett strukturellt interfederalt samarbetsnätverk för praktisk rådgivning och vägledning och lärande av bästa praxis. Detta kan motverka fragmenteringen av relevant kunskap och kapacitet som behövs för att utveckla eller upphandla AI- och ADM-system som redan i sin utformning omfattar skydd av grundläggande rättigheter, eller åtminstone möjliggör övervakning av effekterna under systemets hela livscykel. I ett sådant nätverk bör det ingå experter på hur man motverkar diskriminering och experter på grundläggande rättigheter och skydd av personuppgifter.
- ▶ **Säkerställa processuella garantier för att göra prövning mer effektiv i fall som rör AI- eller ADM-system, inbegripet att ändra lagstiftningen mot diskriminering vid behov:** Införa omvänd bevisbörda med ett antagande om algoritmisk bias som hjälper individer som annars står inför nästan omöjliga bevishinder. Vid behov bör lagstiftningen mot diskriminering ändras för att stödja detta antagande. Samtidigt bör det bli lättare för allmännyttiga organisationer att inleda kollektiva prövningsförfaranden i sammanhang som

rekrytering, hälso- och sjukvård och social trygghet, vilket är vanligt när det gäller konsumentinriktade regleringar och klagomålsförfaranden.

- ▶ **Genomföra och publicera sammanfattningar av konsekvensbedömningar av grundläggande rättigheter:** Införa ett harmoniserat tillvägagångssätt för konsekvensbedömningar avseende grundläggande rättigheter (FRIA). Dessa bör tillämpas i alla organisationer inom den offentliga sektorn och inbegripa perspektiv från berörda grupper och likabehandlingsorgan. Dessutom bör sammanfattningarna av FRIA offentliggöras.
- ▶ **Inkludera motverkan av diskriminering i offentlig upphandling av AI- och ADM-system:** Säkerställa att krav på att motverka diskriminering integreras i offentlig upphandling av AI och ADM för att förbättra due diligence.
- ▶ **Öka allmänhetens medvetenhet om AI-användning och öka kunskapen om AI:** Kampanjer och insatser för att öka kunskapen om AI, bland annat inom utbildning, bör införas i syfte att öka allmänhetens medvetenhet om användningen av AI i offentliga tjänster.
- ▶ **Garantera likabehandlingsorganens och tillsynsmyndigheters tillgång till tillräckliga offentliga resurser och expertis när det gäller AI-användning:** Dessa är nödvändiga för att likabehandlingsorgan och tillsynsmyndigheter ska kunna fullgöra sitt uppdrag enligt AI-förordningen och lagstiftningen mot diskriminering.
- ▶ **Utbilda personal inom den offentliga sektorn om AI:s och ADM:s inverkan på de grundläggande rättigheterna:** Utbildningen måste gå utöver etik och ta upp AI:s påverkan på de grundläggande rättigheterna, med särskild uppmärksamhet på risker för diskriminering.
- ▶ **Stärka samarbetet mellan likabehandlingsorgan, andra offentliga institutioner och det civila samhället:** Förstärka och institutionalisera samarbete, till exempel genom "praktikgemenskaper" som sänker trösklarna för drabbade individer att söka stöd.

Sammanfattning – Användning av AI i finsk offentlig förvaltning och policyrekommenda- tioner om algoritmisk diskriminering



Sammanfattningen är baserad på rapporten om rådande förhållanden och analys av brister, där den finska delen författades av Emeline Banzuzi.³³

Etablerad användning av automatiserat beslutsfattande, framväxande användning av artificiell intelligens inom offentlig förvaltning

I Finland har offentliga förvaltningar sedan länge använt sig av system för automatiserat beslutsfattande (ADM) för att på ett effektivt sätt hantera ett stort antal administrativa beslut, särskilt inom områden som migration, polisverksamhet, sysselsättning och rekrytering. Användningen av artificiell intelligens (AI) är däremot fortfarande begränsad, och begränsas mestadels till användning av generativ AI (GenAI) i effektivitetssyfte, till exempel Microsoft Copilot som stöd för att skriva dokument eller anteckningar. Flera offentliga myndigheter utforskar eller testar AI-applikationer, däribland finska migrationsverket (Migri), som överväger att använda AI för översättning från tal till text; den finska polisen, som använder AI för att läsa av registreringsskyltar; och Helsingfors stad, som testar "Helsinki GPT", ett AI-verktyg för automatisk kategorisering och prioritering av kundfeedback och generering av svar.

Rättslig och institutionell ram

EU:s AI-förordning utgör tillsammans med europeisk och finsk lagstiftning mot diskriminering och dataskyddslagstiftning ett rättsligt ramverk för att hantera algoritmisk diskriminering. Finska myndigheter förbereder sig för att anpassa sina

33. Detta är en konfidentiell rapport som granskar hur AI- och ADM-system används inom den offentliga sektorn i Finland, utvärderar AI-relaterade rättsliga och politiska ramverk ur ett jämställdhets- och icke-diskrimineringsperspektiv och ger rekommendationer till myndigheter och beslutsfattare.

rättsliga och institutionella ramverk efter de skyldigheter som AI-förordningen inför. Det finns två pågående lagstiftningsförslag för genomförande av AI-förordningen i Finland. Den första syftar till att exempelvis reglera tillsynen av vissa AI-system och verkställigheten av böter samt att i nationell lagstiftning utse de behöriga myndigheterna och en gemensam kontaktpunkt i enlighet med AI-förordningen. Den andra syftar till att reglera till exempel inrättandet och driften av "AI-sandlådor". Andra institutionella förberedelser omfattar inrättande av samordningsmekanismer och klargörande av tillsynsroller.

I Finland regleras användningen av ADM separat i förvaltningslagen, som endast tillåter användning av ADM under omständigheter där det inte krävs någon bedömning i det enskilda fallet för beslutet eller där en tjänsteman själv avgör de delar som förutsätter ärendespecifikt övervägande. ADM måste baseras på förutbestämda regler av en fysisk person.

Diskriminerings- och jämlikhetslagstiftningen i Finland är relativt omfattande, eftersom den stärker främjandet av jämlikhet snarare än att enbart fokusera på att förhindra diskriminering. Skyldigheten att främja jämlikhet kräver att offentliga myndigheter vidtar aktiva och förutseende åtgärder för att säkerställa likabehandling, och denna skyldighet gäller även utformning, utveckling och användning av AI- och ADM-system. Utöver EU:s regelverk ges ytterligare möjligheter att förebygga och hantera algoritmisk diskriminering i Europarådets ramkonvention om artificiell intelligens och mänskliga rättigheter, demokrati och rättsstatsprincipen, som öppnades för undertecknande i september 2024 och undertecknades av EU, där artikel 10 om jämlikhet och icke-diskriminering förbjuder diskriminering och kräver att stater "upprätthålla åtgärder som syftar till att övervinna ojämlikheter för att uppnå skäliga, riktiga och rättvisa resultat".

Den enda kända domen som berör diskriminering och ADM i Finland är ett mål som inbegriper ett företag som tillhandahåller finansiella tjänster där den finska diskriminerings- och jämställdhetsnämnden ålade företaget att upphöra med användningen av statistisk kreditvärderingsmetod som byggde på förbjudna diskrimineringsgrunder, nämligen bostadsort, ålder och kön.³⁴

Brister i rättsliga och politiska ramverk

Behovet av en standardiserad metod för att bedöma konsekvenser för jämlikhet och icke-diskriminering

Rapporten visar att konsekvensbedömningar avseende jämlikhet och icke-diskriminering inte genomförs systematiskt för AI- eller ADM-system, trots att offentliga myndigheter enligt lag är skyldiga att främja jämlikhet och bedöma konsekvenserna för jämlikhet. Ett nationellt bedömningsramverk som tagits fram inom ramen för den

34. Beslut 216/2017 av diskriminerings- och jämställdhetsnämnden i Finland. Tillgänglig på: www.yvtltk.fi/material/sites/ytalkk/tapausselosteet/zedi26zya/Decision_YVTltk_xx16_2017_Multiple_discrimination_in_the_assessment_of_creditworthiness.pdf, hämtad den 10 november 2025.

finska regeringens analys-, bedömnings- och forskningsverksamhet³⁵ tillhandahåller dock ett befintligt verktyg för bedömning av AI-system. Det behövs en gemensam metod för konsekvensbedömningar avseende jämlikhet och grundläggande rättigheter för att säkerställa att jämlikhetsaspekter integreras på ett konsekvent sätt under hela livscykeln för AI- och ADM-system. Diskrimineringsombudsmannen har uppmanat regeringen att se till att offentliga myndigheter har en effektiv och systematiskt tillämpad metod för att bedöma konsekvenser för jämlikhet, bland annat vid offentlig upphandling.³⁶

Bristfällig medvetenhet om risker för diskriminering

Även om tjänstemän på offentliga myndigheter har fått utbildning och tagit fram etiska riktlinjer hade de gynnats av ytterligare resurser, mer riktad utbildning och en djupare förståelse för algoritmisk diskriminering samt lagar mot diskriminering som en del av skyldigheterna i fråga om grundläggande rättigheter.

Begränsade resurser och brist på människorättsbaserad teknisk expertis hos likabehandlingsorgan och tillsynsmyndigheter

Likabehandlingsorgan och andra myndigheter som ansvarar för tillsyn av grundläggande rättigheter förväntas spela en större roll enligt artikel 77 i AI-förordningen, och få befogenheter att begära tillgång till dokumentation och tester av högrisksystem. Deras förmåga att uppfylla denna utökade roll kan dock hindras av begränsade mänskliga, tekniska och ekonomiska resurser, även om användningen av AI-system ökar och allmänhetens medvetenhet och AI-kunskaper fortfarande är begränsade. Dessutom krävs betydande kunskaper om skydd av grundläggande rättigheter, inklusive identifiering och bedömning av olika former av diskriminering, hos alla aktörer med tillsynsuppgifter enligt AI-förordningen. Resurserna bör också utökas för att säkerställa en effektiv förebyggande mekanism för att skydda jämlikheten innan riskerna för diskriminering materialiseras.

Avsaknad av ett offentligt och harmoniserat register över AI- och ADM-system

Det finns ingen enskild offentlig nationell databas som sammanställer information om användningen av AI- eller ADM-system inom den finska offentliga förvaltningen. Användningen av ADM styrs dock av förvaltningslagens krav på öppenhet: myndigheter måste publicera allmän information om användningen av ADM på sina webbplatser och informera enskilda personer när beslut som rör dem har fattats genom ADM. Även om myndigheterna måste publicera information om sådan användning på sina egna webbplatser är denna information fortfarande fragmenterad.

35. Ojanen A. et al. (2022), "Promoting equality in the use of Artificial Intelligence – An assessment framework for non-discriminatory AI", policyöversikt 2022:25.

36. "Diskrimineringsombudsmannens rekommendationer för regeringsprogrammet 2023–2027" (finskt original *Yhdenvertaisuusvaltuutetun suosituksset hallitusohjelmaan 2023–2027*), som finns tillgänglig på: [https://yhdenvertaisuusvaltuutettu.fi/documents/25249352/143315235/Yhdenvertaisuusvaltuutetun+suositukset+hallitusohjelmaan+\(PDF\).pdf/dd8d4ddf-2ab7-9c25-437d-c6619da62712/Yhdenvertaisuusvaltuutetun+suositukset+hallitusohjelmaan+\(PDF\).pdf?t=1681804769918](https://yhdenvertaisuusvaltuutettu.fi/documents/25249352/143315235/Yhdenvertaisuusvaltuutetun+suositukset+hallitusohjelmaan+(PDF).pdf/dd8d4ddf-2ab7-9c25-437d-c6619da62712/Yhdenvertaisuusvaltuutetun+suositukset+hallitusohjelmaan+(PDF).pdf?t=1681804769918), hämtad den 10 november 2025.

Fragmenterad tillsyn och institutionell samordning

Tillsynsansvaret är för närvarande fördelat på flera myndigheter som agerar som nationella behöriga myndigheter, t.ex. marknadstillsynsmyndigheter och myndigheter för grundläggande rättigheter. Samtidigt som denna uppdelning säkerställer att tillsynen täcks mellan olika sektorer riskerar den också att leda till inkonsekventa tolkningar och brister i efterlevnaden av skyddet för de grundläggande rättigheterna, särskilt om samarbetsmekanismerna inte är formaliserade och myndigheterna inte har tillräcklig sakkunskap om hur diskriminering motverkas.

Viktiga rekommendationer

- ▶ **Säkerställa tillräcklig medvetenhet om och tillämpning av lagstiftning mot diskriminering, inklusive främjande av likabehandling:** Detta kan uppnås genom att säkerställa tillräcklig sakkunskap om hur diskriminering motverkas och förhindras, tillhandahålla utbildning och förbättra medvetenheten om och förståelsen för dessa skyldigheter bland behöriga myndigheter. Tillräcklig kunskap och vägledning i arbetet mot diskriminering bör också ges till dem som deltar i eller övervakar upphandlingar och tilldelning som rör inköp och användning av AI.
- ▶ **Införa ett standardiserat ramverk för konsekvensbedömningar av likabehandling och grundläggande rättigheter i AI-systemens processer:** Sådana bedömningar bör tillämpas av alla offentliga myndigheter, med deltagande av likabehandlingsorgan, civilsamhället och berörda grupper. Sammanfattningar av utvärderingar bör publiceras för att öka ansvarsskyldigheten.
- ▶ **Öka resursfördelningen till likabehandlingsorgan och tillsynsmyndigheter:** Nationella behöriga myndigheter som ansvarar för verkställandet av AI-förordningen bör ges tillräckliga resurser för att effektivt kunna utföra sina tillsynsuppgifter och på ett meningsfullt sätt delta i den nationella styrningsramen för AI.
- ▶ **Upprätta ett nationellt offentligt register över AI- och ADM-system som används av offentliga myndigheter, som komplement till den EU-omfattande databas som föreskrivs i AI-förordningen:** Ett sådant register bör öka transparensen, stödja tillsynen och hjälpa likabehandlingsorgan och allmänheten att övervaka risker för diskriminering.
- ▶ **Standardisera mekanismerna för samordning och samarbete mellan behöriga myndigheter, bland annat när det gäller skydd och främjande av jämlikhet:** Detta bör innefatta att förse behöriga myndigheter med tillräcklig expertis inom likabehandling och motverkande av diskriminering för att säkerställa ett enhetligt verkställande av AI-förordningen. Det är viktigt att stärka samarbetet mellan Diskrimineringsombudsmannen och Dataskyddsombudsmannen när det gäller algoritmisk diskriminering, eftersom fall inom detta område kan överlappa rätten till icke-diskriminering och den registrerades rättigheter.

Sammanfattning – Användning av AI i portugisisk offentlig förvaltning och policy- rekommendationer om algoritmisk diskriminering



Sammanfattningen är baserad på rapporten om rådande förhållanden och analys av brister, där den portugisiska delen författades av Eduardo dos Santos.³⁷

Användning av artificiell intelligens inom offentlig förvaltning

I Portugal är användningen av artificiell intelligens (AI) i sin början, men ökar gradvis genom pilotinitiativ och digitaliseringsinsatser. Den nationella portalen för offentlig upphandling listar t.ex. 119 AI-relaterade kontrakt sedan 2009, varav nästan hälften tilldelas enbart under 2023 och 2024.³⁸ AI tillämpas inom sektorer som rättsväsende och sjukvård och flera offentliga myndigheter har lanserat AI-chattbottar för virtuell hjälp med olika tjänster.

Rättslig och institutionell ram

EU:s AI-förordning utgör tillsammans med europeisk och portugisisk lagstiftning mot diskriminering och dataskyddslagstiftning ett rättsligt ramverk för att hantera algoritmisk diskriminering. Utöver EU:s regelverk ges ytterligare möjligheter att förebygga och hantera algoritmisk diskriminering i Europarådets ramkonvention om artificiell intelligens och mänskliga rättigheter, demokrati och rättsstatsprincipen, som öppnades för undertecknande i september 2024, där artikel 10 om jämlikhet och icke-diskriminering förbjuder diskriminering och kräver att stater "upprätthålla åtgärder som syftar till att övervinna ojämlikheter för att uppnå skäliga, riktiga och rättvisa resultat".

37. Detta är en konfidentiell rapport som granskar hur AI- och ADM-system används inom den offentliga sektorn i Portugal, utvärderar AI-relaterade rättsliga och politiska ramverk ur ett jämställdhets- och icke-diskrimineringsperspektiv och ger rekommendationer till myndigheter och beslutsfattare.

38. Portal Base, som finns tillgänglig på: www.base.gov.pt, hämtad den 10 november 2025.

I Portugal inträffade ett anmärkningsvärt rättsfall som rörde algoritmiskt beslutsfattande 2021, när det offentlig-privata flygbolaget TAP använde en algoritm för att välja ut upp till 500 anställda för uppsägning baserat på parametrar som produktivitet, frånvaro, erfarenhet, arbetsinsats, kostnad och kvalifikationer. Arbetstagarna kritiserade bristen på transparens, eftersom viktningen av varje kriterium inte offentliggjordes. År 2023 förklarade Lissabons arbetsdomstol, i ett fall med 23 anställda, att den kollektiva uppsägningen var olaglig.

Den portugisiska regeringen har antagit flera strategiska instrument för att främja digitalisering och ansvarsfull användning av AI till 2030, t.ex. den nationella strategin för artificiell intelligens³⁹ som publicerades 2019 som en del av det nationella initiativet för digital kompetens (INCoDe.2030) och den nationella digitala strategin.⁴⁰ I rapporten om rådande förhållanden och analys av brister konstateras dock att det stora antalet strategier, planer och agendor under årens lopp kan vara kontra-produktivt om de inte samordnas och integreras på rätt sätt.

År 2024 inrättade Portugal rådet för digital offentlig förvaltning med en särskild teknisk arbetsgrupp för artificiell intelligens med målet att övervaka genomförandet av AI-förordningen. Den tidigare myndigheten för administrativ modernisering, som omstrukturerades till myndigheten för teknisk reform av staten 2025, har också en central roll. De säkerställer ansvarsfull användning av digitala system och AI-system i offentlig förvaltning. Offentliga organ som avser att köpa in varor eller tjänster inom detta område, inklusive AI-system, bör informera myndigheten och kan bli föremål för dess bedömning. Myndigheten har tagit fram en vägledning om AI i offentlig förvaltning och ett riskbedömningsverktyg som omfattar vissa frågor om etik, jämlikhet och inkludering.

Brister i rättsliga och politiska ramverk

Begränsad transparens och avsaknad av ett nationellt register över AI och automatiserat beslutsfattande

Det finns inget omfattande offentligt register över de system för AI eller automatiserat beslutsfattande (ADM) som används inom den portugisiska offentliga förvaltningen. För närvarande finns det ingen offentlig mekanism som gör det möjligt för medborgare eller tillsynsorgan att identifiera vilka AI- eller ADM-system som används av offentliga organ. Denna brist på transparens kan göra det svårt att bedöma de potentiella diskrimineringsriskerna och skadorna från sådana system.

Organ för grundläggande rättigheter och likabehandling utesluts från AI-förordningens styrningsram

Portugals förteckning över utsedda myndigheter enligt artikel 77 i AI-förordningen omfattar totalt 14 organ: de allmänna inspektionsmyndigheterna för finans, försvar,

39. National Artificial Intelligence Strategy, som finns tillgänglig på: www.incode2030.gov.pt/en/estrategia-nacional-de-inteligencia-artificial-en, hämtad den 10 november 2025.

40. Portugal Digital Strategy – The key to simplification, som finns tillgänglig på: <https://digital.gov.pt/documentos/portugal-digital-strategy>, hämtad den 10 november 2025.

rättsväsende, intern administration och utbildning, samt sektoriella tillsynsmyndigheter som nationella kommunikationsmyndigheten, myndigheten för livsmedels-säkerhet och ekonomisk säkerhet, och myndigheten för arbetsmiljö. Inga institutioner för grundläggande rättigheter, likabehandlingsorgan eller ombudspersoner ingår som enheter som "skydd av grundläggande rättigheter, inbegripet rätten till icke-diskriminering" i enlighet med artikel 77. Avsaknaden av sådana organ väcker farhågor om hur potentiella fall av algoritmisk diskriminering kommer att hanteras i enlighet med EU-rätten och nationell rätt, samt hur personer som utsätts för diskriminering konsekvent kan begära rättslig prövning enligt AI-förordningens styrningsramverk.

Fragmenterade tillsyns- och samordningsmekanismer

Ansvar för AI-styrning är fördelat över flera myndigheter och det finns inga strukturerade kommunikationskanaler mellan dessa organ och institutioner med mandat inom mänskliga rättigheter eller jämlikhet, till exempel ombudspersonen eller likabehandlingsorgan som kommissionen för medborgarskap och jämställdhet (CIG). Avsaknaden av en samarbetsmekanism gör det svårt att identifiera och åtgärda eventuella överlappningar eller brister i tillsynen, särskilt när det gäller grundläggande rättigheter och skyldigheter inom icke-diskriminering.

Begränsad medvetenhet, resurser och teknisk kapacitet

I Portugal är hänsyn till jämlikhet och mänskliga rättigheter ännu inte systematiskt integrerade i utformningen av AI eller i upphandlingsprocesser, och inte heller i användningen av AI i den offentliga förvaltningen. Likabehandlingsorgan och andra behöriga institutioner har begränsade möjligheter att använda tillräcklig teknisk expertis inom AI och saknar de resurser som krävs för att delta i tekniska utvärderingar av AI-system för att stödja individer som kan påverkas av diskriminering. Kommissionen för medborgarskap och jämställdhet och andra relevanta organ står inför resursbegränsningar som hämmar deras förmåga att ta itu med nya utmaningar, liksom begränsade befogenheter (det vill säga avsaknad av utrednings- och processuella mandat), vilket begränsar deras förmåga att ta itu med frågor om algoritmisk diskriminering.

Viktiga rekommendationer

- **Upprätta ett offentligt register över AI- och ADM-system som används i offentlig förvaltning:** Att införa en skyldighet för offentliga myndigheter att rapportera och dokumentera AI-system i ett centraliserat register hade ökat transparensen, underlättat tillsyns- och likabehandlingsorganens översyn och ökat allmänhetens förtroende.
- **Stärka likabehandlingsorganens utrednings- och verkställighetsbefogenheter och inkludera organ för grundläggande rättigheter och likabehandling som nationella organ enligt artikel 77:** Mot bakgrund av införlivandet av EU:s direktiv om standarder för likabehandlingsorgan,⁴¹ hade detta stärkt dessa institutioner och möjliggjort proaktiv övervakning, inblandning i AI-styrning

41. Direktiv EU 2024/1499 och direktiv EU 2024/1500.

och effektiv efterlevnad av skyldigheter att motverka diskriminering enligt både europeisk och nationell lagstiftning.

- ▶ **Fram till dess att organ för grundläggande rättigheter och likabehandling har utsetts till organ enligt artikel 77, klargöra institutionernas mandat och stärka samordningen mellan myndigheter:** Genom att fastställa tydliga ansvarsområden och formella samarbetsmekanismer mellan likabehandlings-, dataskydds- och konsumentskyddsmyndigheter finns det möjlighet att säkra ett enhetligt informationsutbyte och verkställande av AI-förordningen.
- ▶ **Skapa ett harmoniserat ramverk för att bedöma konsekvenserna för jämlikhet, jämställdhet och grundläggande rättigheter, med utgångspunkt i de nationella AI-relaterade strategierna, vägledning från Agency for Administrative Modernisation och AI-förordningen:** Detta ramverk skulle utveckla ett standardiserat nationellt tillvägagångssätt för att vägleda offentliga myndigheter i utvärderingen av riskerna med algoritmisk diskriminering och säkerställa en konsekvent tillämpning av skyddsåtgärder i den offentliga sektorn.
- ▶ **Tillhandahålla program för utbildning och kapacitetsuppbyggnad om grundläggande rättigheter för offentliga tjänstemän och behöriga myndigheter enligt AI-förordningen:** Riktade utbildnings- och kapacitetsuppbyggnadsinsatser bör genomföras för att ge myndigheter möjlighet att identifiera och åtgärda algoritmisk diskriminering, inklusive könsbaserad diskriminering.
- ▶ **Införliva principer om att motverka diskriminering och främja jämställdhet på ett mer framträdande sätt i offentlig upphandling av AI- och ADM-system:** Detta inkluderar fortsatt övervakning av AI-system med regelbundna revisioner medan de utvecklas och används.
- ▶ **Främja AI-kunskap och bedriva kampanjer för att öka allmänhetens medvetenhet:** Det måste tas fram medvetenhetskampanjer och utbildningsmaterial för att informera allmänheten om hur AI används i offentlig förvaltning, dess potentiella effekter och tillgängliga prövningsmekanismer. Stärkta AI-kunskaper hade ökat transparensen och ansvarstagandet i användningen av dessa system.
- ▶ **Säkerställa tillräcklig finansiering och teknisk AI-expertis för likabehandlingsorgan:** Likabehandlingsorganen måste få tillräcklig finansiering och specialiserad teknisk AI-expertis för att ta itu med algoritmisk diskriminering.
- ▶ **Utforska hur regeringens pågående initiativ för att utveckla en ny generation av flerkanaliga, människocentrerade offentliga tjänster under plattformen gov.pt gör det möjligt för medborgare att lämna in klagomål och anmärkningar gällande AI-driven diskriminering.**

Referenser

Access Now (2024), *Joint statement – A dangerous precedent: how the EU AI Act fails migrants and people on the move*, available at: www.accessnow.org/press-release/joint-statement-ai-act-fails-migrants-and-people-on-the-move, accessed 10 November 2025.

Algorithm Audit (2024a), *Preventing prejudice. Recommendations for risk profiling in the College Grant Control process: a quantitative and qualitative analysis*, available at: https://algorithmaudit.eu/algoprudence/cases/aa202401_preventing-prejudice/, accessed 10 November 2025.

Algorithm Audit (2024b), *DUO control process biased towards students with a non-European migration background*, available at: https://algorithmaudit.eu/events/press_room/#DUO_CBS, accessed 15 November 2025.

Algorithm Watch (2018), *Finnish credit score ruling raises questions about discrimination and how to avoid it*, available at: [Finnish Credit Score Ruling raises Questions about Discrimination and how to avoid it - AlgorithmWatch](#), accessed 15 November 2025.

Allhutter D. et al. (2020), "Algorithmic profiling of job seekers in Austria: how austerity politics are made effective", *Frontiers in Big Data*, Vol. 3, No. 5.

Amnesty International (2021), *Dutch childcare benefit scandal an urgent wake-up call to ban racist algorithms*, available at: www.amnesty.org/en/latest/news/2021/10/xenophobic-machines-dutch-child-benefit-scandal, accessed 10 November 2025.

Conseil d'État (2024), *Requête introductive d'instance*, 15 October.

Council of Europe (2007), *Convention 201 on the Protection of Children against Sexual Exploitation and Sexual Abuse* (Lanzarote, 25 October 2007), available at: <https://rm.coe.int/1680084822>, accessed 10 November 2025.

Council of Europe (2010), *Recommendation CM/Rec(2010)5 of the Committee of Ministers on measures to combat discrimination on grounds of sexual orientation or gender identity*, available at: <https://search.coe.int/cm?i=09000016805cf40a>, accessed 10 November 2025.

Council of Europe (2013), *Declaration by the Committee of Ministers on risks to fundamental rights stemming from digital tracking and other surveillance technologies*, available at: <https://search.coe.int/cm?i=09000016805c8011>, accessed 10 November 2025.

Council of Europe (2018), *Convention 108+ for the protection of individuals with regard to the processing of personal data*, available at: <https://rm.coe.int/convention-108-convention-for-the-protection-of-individuals-with-regar/16808b36f1>, accessed 10 November 2025.

Council of Europe (2019a), *Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes*, Decl(13/02/2019)1, available at: <https://search.coe.int/cm?i=090000168092dd4b>, accessed 10 November 2025.

Council of Europe (2019b), Recommendation CM/Rec(2019)1 of the Committee of Ministers to member States on preventing and combating sexism, available at: <https://rm.coe.int/168093b26a>, accessed 10 November 2025.

Council of Europe (2019c), "Guidelines on artificial intelligence and data protection", Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data (Convention 108), available at: <https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8>, accessed 10 November 2025.

Council of Europe (2020), Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, available at: <https://search.coe.int/cm?i=09000016809e1154>, accessed 10 November 2025.

Council of Europe (2021a), *General Recommendation No. 1 on the digital dimension of violence against women*, Group of Experts on Action against Violence against Women and Domestic Violence (GREVIO), available at: <https://edoc.coe.int/fr/violence-l-gard-des-femmes/10643-grevio-general-recommendation-no-1-on-the-digital-dimension-of-violence-against-women.html>, accessed 10 November 2025.

Council of Europe (2021b), "Content moderation: best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation", Guidance Note, Steering Committee for Media and Information Society (CDMSI), available at: <https://rm.coe.int/content-moderation-en/1680a2cc18>, accessed 10 November 2025.

Council of Europe (2021c), "Guidelines on upholding equality and protecting against discrimination and hate during the Covid-19 pandemic and similar crises in the future", Committee of Ministers, CM(2021)37-add1 rev, available at: <https://edoc.coe.int/en/living-together-diversity-and-freedom-in-europe/9745-guidelines-of-the-committee-of-ministers-of-the-council-of-europe-on-upholding-equality-and-protecting-against-discrimination-and-hate-during-the-covid-19-pandemic-and-similar-crises-in-the-future.html>, accessed 10 November 2025.

Council of Europe (2021d), "Guidelines on facial recognition", Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data (Convention 108), available at: <https://edoc.coe.int/en/artificial-intelligence/9753-guidelines-on-facial-recognition.html>, accessed 10 November 2025.

Council of Europe (2022a), "Outline of Huderia risk and impact assessment methodology", Committee on Artificial Intelligence, available at: <https://rm.coe.int/cai-bu-2022-03-outline-of-huderia-risk-and-impact-assessment-methodolo/1680a81e14>, accessed 10 November 2025.

Council of Europe (2022b), Recommendation CM/Rec(2022)13 of the Committee of Ministers to member States on the impacts of digital technologies on freedom of expression, available at: <https://search.coe.int/cm?i=0900001680a61729>, accessed 10 November 2025.

Council of Europe (2022c), Recommendation CM/Rec(2022)17 of the Committee of Ministers to member States on protecting the rights of migrant, refugee and asylum-seeking women and girls, available at: <https://rm.coe.int/>

[prems-092222-gbr-2573-recommandation-cm-rec-2022-17-a5-bat-web-1-1680a6ef9a#:~:text=Member%20States%20should%20ensure%20that,the%20same%20conditions%20as%20nationals](https://rm.coe.int/prems-092222-gbr-2573-recommandation-cm-rec-2022-17-a5-bat-web-1-1680a6ef9a#:~:text=Member%20States%20should%20ensure%20that,the%20same%20conditions%20as%20nationals), accessed 10 November 2025.

Council of Europe (2022d), Recommendation CM/Rec(2022)16 of the Committee of Ministers on combating hate speech, available at: <https://search.coe.int/cm?i=0900001680a67955>, accessed 10 November 2025.

Council of Europe, Recommendation CM/Rec(2024)4 of the Committee of Ministers on combating hate crime, available at: <https://search.coe.int/cm?i=0900001680af9736>, accessed 10 November 2025.

Council of Europe (2024a), Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, Vilnius.

Council of Europe (2024b), “Methodology for the risk and impact assessment of artificial intelligence systems from the point of view of human rights, democracy and the rule of law (Huderia Methodology)”, Committee on Artificial Intelligence, available at: <https://rm.coe.int/cai-2024-16rev2-methodology-for-the-risk-and-impact-assessment-of-arti/1680b2a09f>, accessed 10 November 2025.

Council of Europe (2024c), Gender Equality Strategy 2024-2029, available at: <https://rm.coe.int/prems-073024-gbr-2573-gender-equality-strategy-2024-29-txt-web-a5-2756/1680afc66a>, accessed 10 November 2025.

Council of Europe (2024d), Recommendation CM/Rec(2024)1 of the Committee of Ministers to member States on equality of Roma and Traveller women and girls, available at: <https://search.coe.int/cm?i=0900001680af27e4>, accessed 10 November 2025.

Council of Europe (forthcoming), Draft Recommendation of the Committee of Ministers to member States on the impact of artificial intelligence systems, their potential for promoting equality – including gender equality – and the risks they may cause in relation to non-discrimination, Committee of Experts on Artificial Intelligence, Equality and Discrimination (GEC/ADI-AI).

Council of Europe, Commissioner for Human Rights (2019), “Unboxing AI: 10 steps to protect human rights”, available at: <https://book.coe.int/fr/commissaire-aux-droits-de-l-homme/9664-unboxing-artificial-intelligence-10-steps-to-protect-human-rights.html>, accessed 10 November 2025.

Council of Europe, Commissioner for Human Rights (2023), “Human rights by design: future-proofing human rights protection in the era of AI”, Follow-up Recommendation, available at: <https://book.coe.int/fr/commissaire-aux-droits-de-l-homme/11759-pdf-human-rights-by-design-future-proofing-human-rights-protection-in-the-era-of-ai.html>, accessed 10 November 2025.

Council of Europe, Parliamentary Assembly (2017), Recommendation 2102 (2017) on “Technological convergence, artificial intelligence and human rights”, available at: <https://pace.coe.int/en/files/23726/html>, accessed 10 November 2025.

Council of Europe, Parliamentary Assembly (2020a), Resolution 2343 on “Preventing discrimination caused by the use of artificial intelligence”, available at: <https://pace.coe.int/en/files/28807/html>, accessed 10 November 2025.

Council of Europe, Parliamentary Assembly (2020b), Recommendation 2183 (2020) on “Preventing discrimination caused by the use of artificial intelligence”, available at: <https://pace.coe.int/en/files/28809/html>, accessed 10 November 2025.

Court of Justice of the European Union (2008), C-54/07 *Centrum voor gelijkheid van kansen en voor racismebestrijding contre Firma Feryn NV* EU:C:2008:397.

Court of Justice of the European Union (2022), C-817/19 *Ligue des droits humains ASBL v. Conseil des ministres* EU:C:2022:491.

Court of Justice of the European Union (2023), C-634/21 *SCHUFA Holding* (Scoring) EU:C:2023:957.

Court of Justice of the European Union (2024), C-588/21 P *Public.Resource.Org and Right to Know v. Commission and Others* EU:C:2024:201.

Court of Justice of the European Union (2025), C-203/22 *Dun & Bradstreet Austria* EU:C:2025:117.

Défenseur des Droits (2024), “Algorithms, AI systems and public services: what rights do users have?”, available at: www.defenseurdesdroits.fr/our-last-reports-and-studies-316, accessed 10 November 2025.

Dubois V. (2021), *Contrôler les assistés : Genèses et usages d’un mot d’ordre*, Raisons d’agir, Paris.

European Commission (2020), “White Paper: On artificial intelligence – A European approach to excellence and trust”, COM(2020)65 final.

European Commission (2022a), “Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI liability directive)”, COM/2022/496 final.

European Commission (2022b), “European Declaration on Digital Rights and Principles for the Digital Decade”, COM(2022) 28 final.

European Commission (2025a), “Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)”, available at: <https://ec.europa.eu/newsroom/dae/redirection/document/112367>, accessed 10 November 2025.

European Commission (2025b), *The General-Purpose AI Code of Practice*, available at: <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai>, accessed 10 November 2025.

European Commission (2025c), *AI Board*, <https://digital-strategy.ec.europa.eu/en/policies/ai-board>, accessed 10 November 2025.

European Commission (2025d), “Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)”, C(2025) 924 final.

European Commission against Racism and Intolerance (2000), “ECRI General Policy Recommendation No. 6 on combating the dissemination of racist, xenophobic and antisemitic material via the internet”, available at: <https://>

rm.coe.int/ecri-general-policy-recommendation-no-6-on-combating-the-dissemination/16808b5a8d, accessed 10 November 2025.

European Commission against Racism and Intolerance (2004), “ECRI General Policy Recommendation No. 8 on combating racism while fighting terrorism”, available at: <https://rm.coe.int/ecri-general-policy-recommendation-no-8-on-combating-racism-while-figh/16808b5abc>, accessed 10 November 2025.

European Commission against Racism and Intolerance (2006), “ECRI General Policy Recommendation No. 10 on combating racism and racial discrimination in and through school education”, available at: <https://rm.coe.int/ecri-general-policy-recommendation-no-10-on-combating-racism-and-racia/16808b5ad5>, accessed 10 November 2025.

European Commission against Racism and Intolerance (2007), “ECRI General Policy Recommendation No. 11 on combating racism and racial discrimination in policing”, available at: <https://rm.coe.int/ecri-general-policy-recommendation-no-11-on-combating-racism-and-racia/16808b5adf>, accessed 10 November 2025.

European Commission against Racism and Intolerance (2012), “ECRI General Policy Recommendation No. 14 on combating racism and racial discrimination in employment”, available at: <https://rm.coe.int/ecri-general-policy-recommendation-no-14-on-combating-racism-and-racia/16808b5afc>, accessed 10 November 2025.

European Commission against Racism and Intolerance (2015), “ECRI General Policy Recommendation No. 15 on combating hate speech”, available at: <https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech/16808b5b01>, accessed 10 November 2025.

European Commission against Racism and Intolerance (2017), “ECRI General Policy Recommendation No. 7 (revised) on national legislation to combat racism and racial discrimination”, available at: www.coe.int/en/web/european-commission-against-racism-and-intolerance/recommendation-no.7, accessed 10 November 2025.

European Court of Human Rights (2007), Application No. 57325/00, *D.H. and Others v. the Czech Republic* (Grand Chamber, 13 November 2007), paragraph 178.

European Court of Human Rights (2023), Application 11519/20, *Glukhin v. Russia* (Third Section, 2023).

European Digital Rights (2024), *How to fight biometric mass surveillance after the AI Act: a legal and practical guide*, available at: <https://edri.org/our-work/how-to-fight-biometric-mass-surveillance-after-the-ai-act-a-legal-and-practical-guide>, accessed 10 November 2025.

European Migration Network and Organisation for Economic Co-operation and Development (2022), “The use of digitalisation and artificial intelligence in migration management”, available at: www.oecd.org/en/topics/sub-issues/migration-policies-returns-and-attracting-talent/migration-policy-debates-and-data-briefs.html, accessed 10 November 2025.

European Network of National Human Rights Institutions (2024), “Technologies, migration, and human rights: the role of European NHRIs”, ENNHRI scoping paper,

available at: <https://ennhri.org/publications-statements/#artificial-intelligence-publications>, accessed 10 November 2025.

European Union (2016a), Regulation 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.

European Union (2016b), Directive 2016/680 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data.

European Union (2022), Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services.

European Union (2024a), Council Directive 2024/1499 on standards for equality bodies in the field of equal treatment between persons irrespective of their racial or ethnic origin, equal treatment in matters of employment and occupation between persons irrespective of their religion or belief, disability, age or sexual orientation, equal treatment between women and men in matters of social security and in the access to and supply of goods and services.

European Union (2024b), Council Directive 2024/1500 on standards for equality bodies in the field of equal treatment and equal opportunities between women and men in matters of employment and occupation.

European Union (2024c), Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689, accessed 10 November 2025.

Lanzarote Committee (2024), Declaration on protecting children against sexual exploitation and sexual abuse facilitated by emerging technologies, available at: <https://rm.coe.int/declaration-on-protecting-children-against-sexual-exploitation-and-sex/1680b25a78>, accessed 10 November 2025.

La Quadrature du Net (2023), *Notation des allocataires : l'indécence des pratiques de la CAF désormais indéniable*, available at: www.laquadrature.net/2023/11/27/notation-des-allocataires-lindcence-des-pratiques-de-la-caf-desormais-indeniable, accessed 10 November 2025.

La Quadrature du Net (2024), *À France Travail, l'essor du contrôle algorithmique*, 25 June, available at: www.laquadrature.net/2024/06/25/a-france-travail-lessor-du-controle-algorithmique, accessed 10 November 2025.

Malgieri G. and Comandé G. (2017), "Why a right to legibility of automated decision-making exists in the General Data Protection Regulation", *International Data Privacy Law*, Vol. 7, Issue 4, p. 243-65.

McGregor L. and Molnar P. (2023), "Digital border governance: a human rights based approach", University of Essex and the United Nations Human Rights Office of the High Commissioner.

Milano S. and Prunkl C. (2024), "Algorithmic profiling as a source of hermeneutical injustice", *Philosophical Studies*, Vol. 182, pp. 185-203.

Mittelstadt B. (2025), "How to use the Artificial Intelligence Act to investigate AI bias and discrimination: a guide for Equality Bodies", European Network of Equality Bodies (Equinet), Brussels.

National Non-Discrimination and Equality Tribunal of Finland (2017), Decision 216/2017.

Selbsts A. D. and Powles J. (2017) "Meaningful information and the right to explanation", *International Data Privacy Law*, Vol. 7, No. 4, pp. 233-42.

UN Special Rapporteur (2021), *Racial and Xenophobic discrimination and the use of digital technologies in border and immigration enforcement – Report of the Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance*, Achiume E. Tendayi, UN Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, A/HRC/48/76.

Wachter S., Mittelstadt B. and Floridi L. (2017), "Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation", *International Data Privacy Law*, Vol. 7, Issue 2, pp. 76-99.

Xenidis R. (2021), "Tuning EU equality law to algorithmic discrimination: three pathways to resilience", *Maastricht Journal of European and Comparative Law*, Vol. 27, Issue 6, pp. 736-58.

Den ökande integreringen av artificiell intelligens (AI) i olika tillämpningar inom ett brett spektrum av sektorer innebär betydande utmaningar för skyddet av grundläggande rättigheter. Bland dessa framträder algoritmisk diskriminering som ett särskilt angeläget problem. Empirisk forskning har visat att algoritmisk partiskhet inte bara återspeglar utan också förvärrar befintliga sociala ojämlikheter i stor skala, särskilt inom områden där algoritmisk beslutsfattande är vanligt förekommande – såsom polisarbete, sysselsättning, utbildning och försäkring. År 2024 antog både Europeiska unionen och Europarådet banbrytande rättsliga instrument: EU:s AI-förordning (förordning 2024/1689) och Europarådets ramkonventionen om artificiell intelligens och mänskliga rättigheter, demokrati och rättsstatsprincipen. I denna rapport undersöks hur dessa nya rättsliga instrument bidrar till att stärka skyddet mot algoritmisk diskriminering i Europa och bedöms de luckor som fortfarande påverkar dessa rättsliga ramar. Rapporten syftar också till att spåra och sammanfatta de senaste utvecklingen för att förse relevanta aktörer, såsom likabehandlingsorgan, med verktyg för att anpassa sig till ett föränderligt rättsligt landskap och, där så är relevant, ingripa i pågående utvecklingar.



YHDENVERTAISUUSVALTUUTETTU
NON-DISCRIMINATION OMBUDSMAN



Medlemsstaterna i Europeiska unionen har beslutat att förena sina kunskaper, resurser och öden. Tillsammans har de byggt upp en zon av stabilitet, demokrati och hållbar utveckling samtidigt som de har bevarat kulturell mångfald, tolerans och individers friheter. Europeiska unionen är fast besluten att dela med sig av sina framsteg och sina värderingar till länder och folk utanför sina gränser.

Europarådet är kontinentens ledande organisation för de mänskliga rättigheterna. Det består av 46 medlemsstater, inklusive samtliga medlemmar av Europeiska unionen. Alla medlemsstater i Europarådet har undertecknat den europeiska konventionen om skydd för de mänskliga rättigheterna och de grundläggande friheterna (Europakonventionen), ett fördrag som syftar till att skydda de mänskliga rättigheterna, demokrati och rättsstatsprincipen. Europeiska domstolen för de mänskliga rättigheterna (Europadomstolen) övervakar genomförandet av Europakonventionen i medlemsstaterna.

www.europa.eu

www.coe.int



EUROPEAN UNION

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE