

OIKEUSSUOJA ALGORITMISTA SYRJINTÄÄ VASTAAN EUROOPASSA

Nykyiset puitteet ja
jäljellä olevat puutteet



Yhteistyössä:

Interfederal Centre for Equal Opportunities (Unia), Belgia,
Yhdenvertaisuusvaltuutettu (YVV), Suomi, ja
Commission for Citizenship and Gender Equality (CIG), Portugali

Raphaële Xenidis

Euroopan unionin
yhteisrahoittama



EUROOPAN UNIONI

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE

Euroopan neuvoston
yhteisrahoittama ja toteuttama

OIKEUSSUOJA ALGORITMISTA SYRJINTÄÄ VASTAAN EUROOPASSA

Nykyiset puitteet ja
jäljellä olevat puutteet

Raphaële Xenidis

Raportti on laadittu Euroopan unionin ja Euroopan neuvoston taloudellisella tuella. Sen sisällöstä vastaa yksin kirjoittaja. Tässä raportissa esitettyjen näkemysten ei voida missään tapauksessa katsoa edustavan Euroopan unionin tai Euroopan neuvoston virallista kantaa.

Otteiden (enintään 500 sanaa) jäljentäminen on sallittua kaupallisia tarkoituksia lukuun ottamatta, kunhan tekstin eheys säilyy, otetta ei käytetä asiayhteydestään irrotettuna, ote ei tarjoa epätäydellisiä tietoja eikä lukijaa muutoin johdeta harhaan tekstin luonteen, laajuuden tai sisällön suhteen. Lähdeteksti on aina mainittava seuraavasti: "© Euroopan neuvosto, julkaisuvuosi". Kaikki muut asiakirjan tai sen osan jäljentämistä/kääntämistä koskevat pyynnöt on osoitettava Euroopan neuvoston julkaisuista ja visuaalisesta ilmeestä vastaavalle osastolle (Publications and Visual Identity Division) (F-67075 Strasbourg Cedex tai publishing@coe.int).

Kaikki muu tätä asiakirjaa koskeva viestintä on osoitettava Euroopan neuvoston osallisuus- ja syrjinnäntorjuntaohjelmista vastaavan osaston (Inclusion and Anti-discrimination Programmes Division) vihapuheen, viharikosten ja tekoälyn yksikölle, F-67075 Strasbourg Cedex, Ranska. Sähköposti: anti-discrimination@coe.int

Alkuperäinen englanninkielinen teksti on Euroopan neuvoston laatima. Tämä suomenkielinen käännös on julkaistu Euroopan neuvoston suostumuksella, mutta sen sisällöstä vastaavat yksin kääntäjät. Euroopan neuvoston julkaisujen ja visuaalisen ilmeen osaston toimitusyksikkö ei ole korjannut tämän suomenkielisen julkaisun typografisia tai kieliopillisia virheitä.

Kannen suunnittelu: Publications and Visual Identity Division (DPiV), Euroopan neuvosto
Englanninkielisen julkaisun taitto: Jouve, Paris
Suomenkielisen julkaisun taitto: Vida Design, Pori
Käännös: ALM Translations
Käännöksen oikeudellinen tarkastus: Walther Roslin, 1001 Lakes
Kuvat: Shutterstock

Sisällys

LYHENTEET	7
KIITOKSET	8
JOHDANTO	9
1. OLEMASSA OLEVAT OIKEUDELLISET KEHYKSET	11
1.1. Määritelmät	11
1.2. Algoritminen syrjintä Euroopassa	12
1.3. Tekoälyn ja ADM-järjestelmien käyttö julkishallinnossa ja muilla sovellettavilla sektoreilla	13
1.4. Olemassa olevat oikeudelliset kehykset	17
1.4.1. Tekoälyn ja ADM-järjestelmien hallinnointi	17
1.4.2. Syrjinnän torjunnan välineet	30
1.4.3. Tietosuojavälineet	36
1.4.4. Euroopan ihmisoikeustuomioistuimen ja Euroopan unionin tuomioistuimen oikeuskäytäntö	39
1.5. Keskeiset sidosryhmät	42
2. OIKEUDELLISET PUUTTEET	46
2.1. Perustelut tekoälyn ja ADM-järjestelmien käyttöönotolle	46
2.2. Avoimuuden puute	47
2.3. Oikeussuojan saatavuushaasteet	49
2.4. Nykyisten ja tulevien säännösten täytäntöönpano	51
2.4.1. Instituutioiden välinen yhteistyö	51
2.4.2. Käsitteelliset puutteet	51
2.4.3. Tietojen keräämistä ja käsittelyä koskevat rajoitukset	53
2.5. Tekoälyjärjestelmien ”riskien vähentäminen”	53
2.6. Tekniset standardit, yhdenvertaisuuselimet ja yhdenmukaistamiseen liittyvä kysymys	54
TIIVISTELMÄ – TEKÖÄLYN KÄYTTÖ BELGIAN JULKISHALLINNOSSA JA ALGORITMISTA SYRJINTÄÄ KOSKEVAT POLIITTISET SUOSITUKSET	59
TIIVISTELMÄ – TEKÖÄLYN KÄYTTÖ SUOMEN JULKISHALLINNOSSA JA ALGORITMISTA SYRJINTÄÄ KOSKEVAT POLIITTISET SUOSITUKSET	63
TIIVISTELMÄ – TEKÖÄLYN KÄYTTÖ PORTUGALIN JULKISHALLINNOSSA JA ALGORITMISTA SYRJINTÄÄ KOSKEVAT POLIITTISET SUOSITUKSET	68
LÄHDEVIITTEET	73

Lyhenteet

AI	tekoäly
ADM	automaattinen päätöksenteko
CSO:t	kansalaisjärjestöt
CAI	Euroopan neuvoston tekoälykomitea
CEN	Euroopan standardointikomitea
CENELEC	Euroopan sähkötekniikan standardointikomitea
FRIA	perusoikeusvaikutusten arviointi
GDPR	yleinen tietosuoja-asetus
GPAI	yleiskäyttöinen tekoäly
HUDERIA	menetelmä, jolla arvioidaan tekoälyjärjestelmistä aiheutuvia ihmisoikeuksiin, demokratiaan ja oikeusvaltioperiaatteeseen kohdistuvia riskejä ja vaikutuksia
SEU	sopimus Euroopan unionista
SEUT	sopimus Euroopan unionin toiminnasta

Kiitokset

Tämä raportti laadittiin hankkeen ”Upholding equality and non-discrimination by Equality bodies regarding the use of artificial intelligence in public administrations” yhteydessä. Hankkeen rahoittajina toimivat Euroopan unioni teknisen tuen välineen kautta sekä Euroopan neuvosto. Hankkeen toteuttamisesta vastaa Euroopan neuvosto yhteistyössä Euroopan komission, Belgian yhdenvertaisuustyöhön erikoistuneen instituution (Interfederal Centre for Equal Opportunities, Unia), Suomen yhdenvertaisuusvaltuutetun (YVV) ja Portugalin kansalaisuuksien ja sukupuolten yhdenvertaisuutta käsittelevän komission (Commission for Citizenship and Gender Equality, CIG) kanssa.

Raportin laati Raphaële Xenidis (eurooppaoikeuden apulaisprofessori, Sciences Po -yliopisto, oikeustieteellinen tiedekunta, Ranska). Kansallisten kontekstien tiivistelmät perustuvat Ine van Zeelandin (Euroopan neuvoston kansallinen konsultti) työhön Belgian osiossa, Emeline Banzuzin (Euroopan neuvoston kansallinen konsultti) työhön Suomen osiossa ja Eduardo dos Santosin (Euroopan neuvoston kansallinen konsultti) työhön Portugalin osiossa.

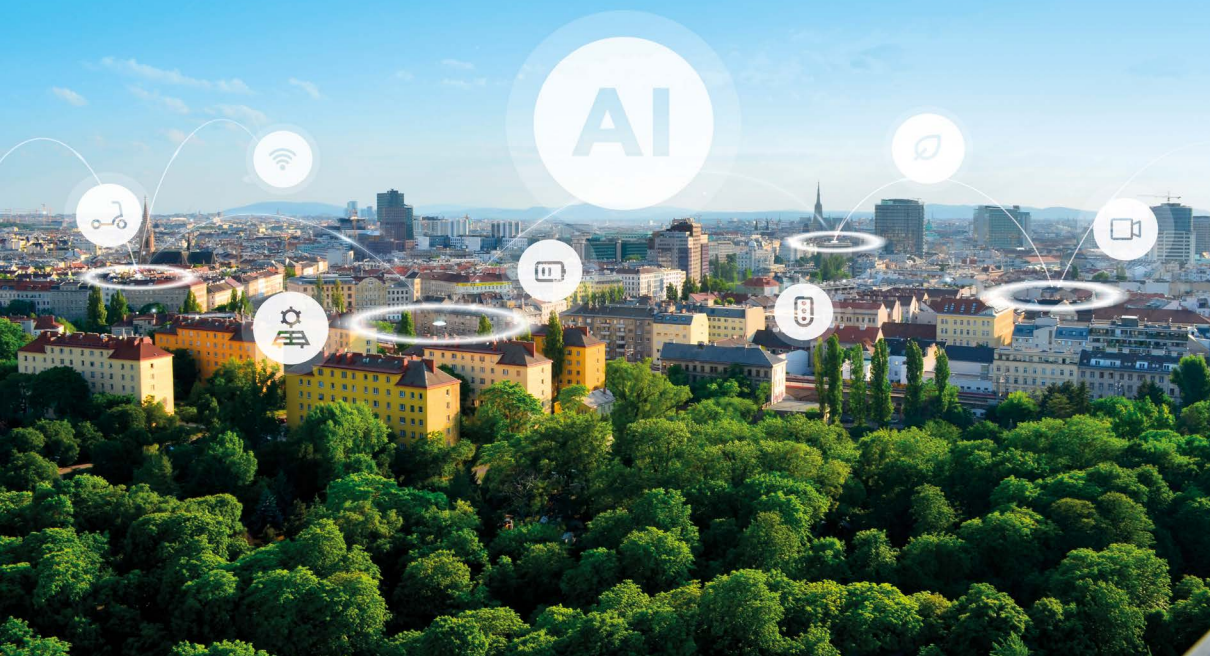
Euroopan neuvosto haluaa kiittää hankkeeseen osallistuneita instituutioita ja Euroopan komissiota jatkuvasta sitoutumisesta raportin laatimisprosessin ajan sekä erityisesti seuraavia henkilöitä: Nele Roekens, Nadine Brauns ja Thelma Raes (Belgian Interfederal Centre for Equal Opportunities, Unia); Tiina Valonen ja Ville Rantala (Suomen yhdenvertaisuusvaltuutettu, YVV); Carla Peixe, Ana Martinho Fernandes, Alexandra Andrade ja Susana Miguel (Portugalin Commission for Citizenship and Gender Equality, CIG) ja Massimiliano Santini (Euroopan komission uudistus- ja investointityöryhmän pääsihteeristö) sekä Menno Ettema, Sara Haapalainen, Ayça Dibekoğlu ja Delfine Gaillard (Euroopan neuvoston syrjinnän ehkäisemistä käsittelevä osasto).

Raportissa hyödynnettiin myös vuoden 2024 lopussa ja vuoden 2025 alussa haastateltujen Belgian, Suomen ja Portugalin julkisten ja yksityisten tahojen kansallisten asiantuntijoiden panosta. Heidän näkemyksensä rikastuttivat raportin kansallisia osioita.



YHDENVERTAISUUSVALTUUTETTU
NON-DISCRIMINATION OMBUDSMAN





Johdanto

Tekoälyn (AI) kasvava integroituminen moniin eri käyttösovelluksiin useilla eri sektoreilla asettaa merkittäviä haasteita perusoikeuksien suojelulle. Näistä haasteista algoritmisen syrjintä on erityisen kiireellinen huolenaihe. Empiiriset tutkimukset ovat osoittaneet, että algoritmiset vinoumat eivät ainoastaan vastaa vaan myös pahentavat laajamittaisesti olemassa olevaa sosiaalista eriarvoisuutta. Tämä koskee erityisesti aloja, joilla algoritmisen päätöksenteko on yleistä, kuten poliisitoimintaa, työllisyyttä, koulutusta ja vakuutuspalveluita. Yhdenvertaisuus- ja tietosuojalainsäädännön yhdistelmä toimii pitkään ensisijaisena oikeusperustana algoritmisen syrjinnän käsittelylle Euroopassa.

Viimeaikaiset oikeudelliset, institutionaaliset ja poliittiset kehitykset ovat kuitenkin muuttaneet merkittävästi tekoälyn hallinnollisia ympäristöjä Euroopassa. Vuonna 2024 sekä Euroopan neuvosto että Euroopan unioni hyväksyivät merkittäviä oikeudellisia välineitä: tekoälyä ja ihmisoikeuksia, demokratiaa ja oikeusvaltioperiaatetta koskevan Euroopan neuvoston puiteyleissopimuksen (tekoälyä koskeva puiteyleissopimus) sekä Euroopan parlamentin ja neuvoston asetuksen (EU) 2024/1689, annettu 13 päivänä kesäkuuta 2024, tekoälyä koskevista yhdenmukaistetuista säännöistä (EU:n tekoälyasetus). Nämä säädökset luovat kattavat sääntelykehikset, joiden tarkoituksena on turvata perusoikeudet – mukaan lukien oikeus syrjimättömyyteen – tekoälyn ja automaattisen päätöksenteon (ADM) järjestelmien kontekstissa. Näiden puitteiden

hyväksymisen odotetaan aiheuttavan merkittäviä oikeudellisia ja institutionaalisia vaikutuksia kansallisella tasolla erityisesti sen vuoksi, että vuoteen 2026 mennessä jäsenvaltioiden on yhdenmukaistettava kansallinen lainsäädäntönsä ja hallinto-rakenteensa EU:n tekoälyasetuksessa säädettyjen velvoitteiden kanssa.

Tekoälyasetuksessa korostetaan, että tekoälyteknologioita ja niiden sääntelykehyksiä on kehitettävä Euroopan unionista tehdyn sopimuksen (SEU) 2 artiklan, sopimusten takaamien perusoikeuksien ja vapauksien sekä SEU:n 6 artiklan mukaisen perusoikeus-kirjan kuvaamien EU:n arvojen mukaisesti. Asetuksessa todetaan nimenomaisesti, että tekoälyn tulee olla ihmiskeskeistä teknologiaa. Vastaavasti Euroopan neuvoston tekoälyä koskevassa puiteyleissopimuksessa vahvistetaan tavoite varmistaa, että kaikki tekoälyn elinkaaren aikana toteutettavat toimet ovat täysin ihmisoikeuksien, demokratian ja oikeusvaltioperiaatteiden mukaisia.

Tässä raportissa tarkastellaan, miten nämä uudet säädökset auttavat vahvistamaan suojaustoimia algoritmista syrjintää vastaan Euroopassa, ja arvioidaan puutteita, jotka yhä vaikuttavat näihin oikeudellisiin kehyksiin.

Samanaikaisesti EU:n kahdella uudella yhdenvertaisuuselimillä koskevalla direktii-villä pyritään vahvistamaan näiden instituutioiden roolia ja valmiuksia Euroopassa (Euroopan unioni, 2024a ja 2024b). Yhdenvertaisuuselimillä sekä kansallisilla ihmis-oikeusinstituutioilla ja oikeusasiamiehillä odotetaan olevan keskeinen rooli teko-älyn ja ADM-järjestelmien hallinnoinnissa ja erityisesti perusoikeusnormien, kuten yhdenvertaisuusperiaatteen, noudattamisen varmistamisessa. Siksi tässä raportissa pyritään jäljittämään ja kokoamaan yhteen nämä kehitykset, jotta asianomaiset toimijat, kuten yhdenvertaisuuselimet, voivat saada käyttöönsä välineitä sopeutua muuttuvaan oikeusympäristöön ja tarvittaessa puuttua meneillään oleviin kehityksiin.

Raportti koostuu kahdesta pääosasta:

- ▶ olemassa olevien oikeudellisten kehysten arviointi
- ▶ jäljellä olevien puutteiden arviointi.

Ensimmäisessä osiossa esitetään määritelmiä (1) ja esimerkkejä tekoälyn ja ADM-järjestelmien käyttöönotosta ja käytöstä julkishallinnossa ja muilla sovellettavilla aloilla Euroopan tasolla (2). Kun on selitetty, miten algoritmista syrjintää esiintyy (3), kartoitetaan olemassa olevat syrjinnän torjuntaan, yhdenvertaisuuteen ja tekoälyyn liittyvät oikeudelliset, menettelylliset ja hallinnolliset puitteet Euroopassa (4). Osion lopuksi tarkastellaan asianomaisten sidosryhmien nykyistä ja mahdollista roolia algoritmisessa syrjinnässä (5).

Toisessa osiossa kartoitetaan näihin puitteisiin ja niiden täytäntöönpanoon liittyviä oikeudellisia puutteita. Tämän raportin soveltamisala rajoittuu ensisijaisesti euroop-palaisten julkishallintojen käyttämiin tekoäly- ja ADM-järjestelmiin ja toissijaisesti yksityisellä sektorilla, kuten suurissa vuokratyövoimaa välittävissä yrityksissä, käy-tettäviin järjestelmiin.



1. Olemassa olevat oikeudelliset kehykset

1.1. Määritelmät

Tässä raportissa termejä ”tekoäly” ja ”ADM-järjestelmät” käytetään kuvaamaan sellaisten oppimiseen ja sääntöihin perustuvien järjestelmien toimintaa, jotka voivat sekä ylläpitää että vahvistaa syrjintää. Euroopan neuvoston tekoälyä koskevan puiteyleissopimuksen 2 artiklassa tekojärjestelmällä tarkoitetaan ”konepohjaista järjestelmää, joka eksplisiittisten tai implisiittisten tavoitteiden saavuttamiseksi päättelee vastaanottamistaan syöttötiedoista, kuinka tuottaa tuotoksia, kuten ennusteita, sisältöä, suosituksia tai päätöksiä, jotka voivat vaikuttaa todellisiin tai virtuaalisiin ympäristöihin.” Siinä tunnustetaan, että ”[e]ri tekoälyjärjestelmillä on eriasteinen kyky toimia itsenäisesti ja mukautua käyttöönoton jälkeen” (Euroopan neuvosto, 2024a). Tämä määritelmä vastaa EU:n tekoälyasetuksen 3 artiklan 1 kohdassa esitettyä määritelmää.¹ Tekoälytekniikoissa

1. Euroopan parlamentin ja neuvoston asetuksen (EU) 2024/1689, annettu 13 päivänä kesäkuuta 2024 tekoälyä koskevista yhdenmukaistetuista säännöistä, 3 artiklan 1 kohdassa tekoälyjärjestelmän määritellään tarkoittavan ”konepohjaista järjestelmää, joka on suunniteltu toimimaan käyttöönoton jälkeen vaihtelevilla autonomian tasoilla ja jossa voi ilmetä mukautuvuutta käyttöönoton jälkeen ja joka päättelee vastaanottamastaan syötteestä eksplisiittisiä tai implisiittisiä tavoitteita varten, miten tuottaa tuotoksia, kuten ennusteita, sisältöä, suosituksia tai päätöksiä, jotka voivat vaikuttaa fyysisiin tai virtuaalisiin ympäristöihin”.

on käytetty erityisesti koneoppimista laajalti niin kutsuttujen tietopisteiden välisten korrelaatioiden tunnistukseen. Tätä käytetään ennusteiden ja suositusten tekemiseen ja esineiden tai ihmisten pisteyttämiseen, luokitteluun tai paremmuusjärjestykseen asettamiseen.² Tätä vastoin tietyiltä ADM-järjestelmiltä puuttuu tekoälyjärjestelmien luokitukseen kuuluva riittävä mukautuvuus ja autonomia, mutta niitä on kuitenkin tärkeää tarkastella tässä raportissa niiden syrjintäpotentiaalin vuoksi. Esimerkiksi sääntöpohjaiset järjestelmät voivat luokitella yksilöitä väärin ja sitten automatisoida päätöksentekoprosesseja tämän luokittelun perusteella, millä voi olla syrjiviä seurauksia. Termi ”ADM-järjestelmät” kattaa monenlaisia päätöksentekojärjestelmiä. Ne vaihtelevat täysin automaattisista järjestelmistä, joihin ei liity lainkaan ihmisen syötteitä, kuten esimerkiksi EU:n yleisen tietosuojasetuksen (GDPR) 22 artiklassa on määritelty, osittain automaattisiin tai päätöksenteon tukijärjestelmiin, joihin sisältyy sekä kone- että ihmissyötteitä ja joita käytetään käytännössä hyvin laajasti.³ Tässä raportissa keskitytään sekä tekoälyyn että ADM-järjestelmiin.

1.2. Algoritminen syrjintä Euroopassa

Algoritmisessa päätöksenteossa syrjintää esiintyy, kun automaattiset järjestelmät asettavat epäoikeudenmukaisesti tietyt yksilöt tai ryhmät epäedulliseen asemaan. Algoritmiseen syrjintään vaikuttavat useat erilaiset avaintekijät.

Ensinnäkin koulutus- ja arviointitiedoissa esiintyvät vinoumat voivat johtaa syrjiviin päätöksiin. Vääristyneet tiedonkeruu- ja tuotantokäytännöt voivat johtaa datajoukkoihin, jotka eivät edusta todellisia tietoja. Jos tekoäly ja datapohjaiset ADM-järjestelmät kuitenkin koulutetaan historiatiedoilla, jotka sisältävät esimerkiksi tiettyjä etnisiä alkuperiä, sukupuolia tai sosioekonomisia asemia koskevia ennakkoluuloja, eriarvoisuutta tai syrjintää, tai ne käsittelevät näitä historiatietoja, järjestelmät usein myös jäljentävät ja/tai voimistavat näitä vinoumia. Esimerkiksi aiemmilla rekrytointitiedoilla koulutettu valinta-algoritmi voi suosia epäoikeudenmukaisesti miespuolisia ehdokkaita jäljentämällä aiempia rekrytointipäätöksiä, joita ovat vääristäneet naisiin tai vähemmistöryhmiin kohdistuvat syrjivät stereotypiat.

Toiseksi kehittäjien suunnitteluvaiheessa tekemät valinnat, kuten kohdeominaisuuksien ja seurausmuuttujien valinta, mallin tyyppi, oikeudenmukaisuusmittarit ja kynnysarvot, voivat sisältää syrjiviä vinoumia. Jos kehittäjät eivät ota huomioon systeemistä eriarvoisuutta, tekoäly- ja ADM-järjestelmät voivat tahattomasti tehdä päätöksiä, jotka vahingoittavat suhteettomasti tiettyjä ryhmiä. Esimerkiksi eurooppalaisissa sosiaalipalveluissa käytetään yhä useammin ADM-järjestelmiä sosiaaliturvatuuksien saajien mahdollisten petosten ennustamiseen. On osoitettu, että ”riskin” kategoria itse ja näiden ADM-järjestelmien operationalisoimat seurausmuuttujat sisältävät

2. Binnsin mukaan "[s]uurin osa koneoppimisen käytöstä julkisella sektorilla on ohjattua. Ohjauksella tarkoitetaan, että oppimisalgoritmilta on näytettävä, mitä päätöksentekijä haluaa ennustaa tai luokitella, toisin kuin ohjaamattomilla menetelmillä, jotka on suunniteltu löytämään piilorakenteita datajoukoista". Katso Binns R. (2020), "Algorithmic decision-making: a guide for lawyers", *Judicial Review*, 25(2), s. 3–4.
3. Näiden erilaisten ADM-järjestelmien taksonomiaa varten katso Palmiotto F. (2024), "When is a decision automated? A taxonomy for a fundamental rights analysis", *German Law Journal*, 25(2), s. 210–36, saatavilla osoitteessa <https://doi.org/10.1017/glj.2023.112>, viitattu 6. marraskuuta 2025.

todellisuudessa köyhyydessä eläviin henkilöihin kohdistuvia stereotypioita, jotka ovat juurtuneet syvälle sosiaalipalvelujen käytäntöihin, muotoihin ja instituutioihin (Dubois 2021). Tämä ongelma kuvastaa sijaismuuttuaan liittyvän syrjinnän laaja-alaisuutta. Tällä viitataan nimenomaan syrjintään, joka perustuu näennäisesti neutraalien mutta kuitenkin näkymättömästi eriarvoisuutta koodaavien sijaismuut-
tujen algoritmiseen käsittelyyn.

Kolmanneksi nykyisin käytössä olevien tekoäly- ja ADM-järjestelmien avoimuuden ja vastuuvollisuuden puute vaikeuttaa algoritmisen syrjinnän valvontaa ja haastamista. Monet tekoäly- ja ADM-järjestelmät, erityisesti monimutkaisiin koneoppimismalleihin perustuvat, toimivat ”mustina laatikoina”, mikä tarkoittaa, etteivät niiden päätöksentekoprosessit ole avoimia. Avoimuuden puute vaikeuttaa sen tunnistamista, milloin ja miten syrjintää esiintyy, ja estää tehokkaan valvonnan ja vastuuvollisuuden soveltamista käytännössä. Esimerkiksi Alankomaissa vuonna 2021 tapahtunut skandaali, jonka aiheutti lastenhoitotukien saajien petosten ennustamiseen käytetty algoritmi, osoitti, kuinka vaikeaa väärin perustein syytettyjen oli ymmärtää, miten ja miksi heidät oli luokiteltu petoksentehtäjäiksi, saati sitten haastaa virheellisesti tehty päätös (Amnesty International 2021).

Lisäksi tekoäly- ja ADM-järjestelmien tarjoajien ja käyttöönottajien suorittama riittämätön testaus ja valvonta sekä julkisella että yksityisellä sektorilla lisää sen todennäköisyyttä, että huomaamaton syrjintää esiintyy. Testaus- ja valvontaprosesseja on käytettävä ennen käyttöönottoa, sen aikana ja sen jälkeen. Asianomaisten kansalaisjärjestöjen (CSO:t), yhdenvertaisuuselinten ja ihmisoikeusinstituutioiden osallistuminen prosesseihin auttaisi varmistamaan näiden menettelyjen tehokkuuden.

Viimeiseksi on todettava, että tekoäly- ja ADM-järjestelmien nopeus ja laajuus uhkaavat levittää syrjintää systeemisellä tasolla. Alan lupaukset tehokkuudesta ja kustannussäästöistä ovat kannustaneet monia organisaatioita, myös julkishallintoja, käyttämään tekoäly- ja ADM-järjestelmiä päätöksenteossa. Tekoäly- ja ADM-järjestelmät ovat kuitenkin väistämättä syrjiviä, jos algoritmisen syrjinnän ehkäisyyn ja lieventämiseen ei investoida riittävästi resursseja esimerkiksi kouluttamalla niiden käyttäjiä, hankkimalla riittävän edustavia tietoja ja käyttämällä asianmukaisia auditointi- ja testausmekanismeja ja vastuuvollisuusprosesseja. Tämän vuoksi on haastettava alan narratiivi, jonka mukaan tekoäly- ja ADM-järjestelmät mahdollistavat merkittävät kustannussäästöt: tekoäly- ja ADM-sovellusten laillisen ja syrjimättömän toiminnan varmistaminen asianmukaisilla suojaustoimilla ja investoinneilla voi todellisuudessa olla kallista.

1.3. Tekoälyn ja ADM-järjestelmien käyttö julkishallinnossa ja muilla sovellettavilla sektoreilla

Tekoälyä ja ADM-järjestelmiä käytetään yhä enemmän sekä yksityisellä että julkisella sektorilla Euroopassa. Kirjallisuuskatsaus osoittaa esimerkkejä käyttösovelluksista lukuisilla aloilla, kuten terveydenhuollossa, rahoitus- ja pankkialalla, liikenteessä ja henkilöstöhallinnossa. Julkishallinnot ottavat asteittain käyttöön tai kokeilevat tekoälyä ja ADM-järjestelmiä esimerkiksi verotuksen, liikenteen, sosiaaliturvan, maahanmuuton sekä oikeuden ja poliisitoiminnan aloilla. Sovelluksia voidaan kehittää itse tai ostaa ulkopuolisilta yksityisiltä yrityksiltä. Tekoäly- ja ADM-järjestelmien pääasiallisia

käyttökohteita näillä aloilla ovat harmittomien hallinnollisten tukitehtävien lisäksi petosten havaitseminen ja valvonta, joihin liittyy suuri syrjinnän riski. Muita laajalle levinneitä sovelluksia ovat chattibotit ja virtuaaliavustajat, jotka voivat myös sisältää syrjiviä piirteitä mutta todennäköisesti vähemmän haitallisin vaikutuksin. Tekoälyn ja ADM-järjestelmien käyttöönotot vaikuttavat kuitenkin erilaisilta: sisäisesti kehitetyt tekoälysovellukset vaikuttavat olevan tällä hetkellä kokeiluvaiheessa, mutta ADM-järjestelmiä käytetään laajemmin.

Tekoälyn ja ADM-järjestelmien käytön kartoittamiseen liittyvät haasteet

Tässä osiossa tehtyyn kartoitukseen liittyy tärkeä varoitus: tekoälyn ja ADM-järjestelmien käytön kartoittaminen julkishallinnossa on vaikea tehtävä. Tutkimuksissa korostetaan, että selkeät ja johdonmukaiset tiedot siitä, miten julkishallinnot käyttävät tekoälyä ja ADM-järjestelmiä, puuttuvat. Joissakin maissa, kuten Suomessa, julkishallinnolla on tiedotusvelvoitteita, joiden vuoksi hallinnon on julkaistava ADM-järjestelmien käyttöä koskevat tiedot verkkosivuillaan ja ilmoitettava niistä päätöksenteon yhteydessä. Tietokantoja, joihin tekoäly- ja ADM-sovellukset kirjataan järjestelmällisesti, ei kuitenkaan ole vielä kehitetty riittävästi.⁴ Näin ollen asiaankuuluvien sidosryhmien, kuten yhdenvertaisuuselinten, kansallisten ihmisoikeusinstituutioiden ja oikeusasiamiesten, on vaikea käyttää toimivaltaansa tutkiakseen, valvoakseen, tunnistaakseen ja haastaakseen algoritmisia syrjintätapauksia sekä tukeakseen uhreja. Nämä esteet vaarantavat syrjimättömyyttä koskevan oikeuden tehokkaan soveltamisen. Myös muut tiedonhankintastrategiat (esimerkiksi tiedonvälityksen vapautta koskevat pyynnöt ja parlamentaariset selvitykset) kohtaavat haasteita ja rajoituksia, eivätkä ne aina anna riittävää tietoa tekoälyn ja ADM-järjestelmien käytöstä julkishallinnossa. Esimerkiksi Ranskassa kansalaisjärjestöjen esittämä tiedonvälityksen vapautta koskeva pyyntö saada selvyyttä perhe-etuuksien myöntämisestä vastaavan sosiaaliviraston (CNAF) käyttämästä riskipisteytysjärjestelmästä johti siihen, että aiempien mallien (eli ei sillä hetkellä käytetyn mallin) lähdekoodi ja käytettyjen muuttujien luettelot julkaistiin rajoitetusti (La Quadrature du Net, 2023). Muita haasteita ovat tekoäly- ja ADM-järjestelmien käytön avoimuuden puute, joka jo lähtökohtaisesti ehkäisee kysymyksiä ja pyyntöjä: jos julkishallinto ei ilmoita tällaisten järjestelmien käytöstä, on vaikea tietää, mistä ylipäätään pitäisi kysyä.

Työllisyys

Raportoituihin käyttötarkoituksiin kuuluvat julkisten työvoimatoimistojen käyttämät profilointityökalut, joiden avulla voidaan ennustaa työllistymis- tai työttömyystodennäköisyyksiä ja joita käytetään tapausten käsittelijöiden apuna tukiresurssien kohdentamisessa työnhakijoille. Osassa tällaisia profilointivälineitä on havaittu syrjintää. Esimerkiksi Itävallassa työvoimatoimiston kehittämän AMS-algoritmi-prototyypin, jonka tarkoituksena oli ennustaa työnhakijoiden työllistymisnäkymiä ja auttaa tapausten käsittelijöitä resurssien kohdistamisessa, osoitettiin syrjivän muun muassa yksinhuoltajaäitejä ja maahanmuuttajataustaisia työnhakijoita (Allhutter et al. 2020). Ranskan työvoimatoimisto France Travail on vuodesta 2018 lähtien käyttänyt

4. Tämä todennäköisesti muuttuu tekoälyasetuksen täytäntöönpanon jälkeen.

vastaavia järjestelmiä petosriskien arviointiin, keskeyttämisten ennakointiin ja tuensaajien työllistettävyyden arviointiin (La Quadrature du Net, 2024). Samankaltaisia syrjintäriskejä, jotka liittyvät sosiaalisesti ja taloudellisesti epävarmoissa oloissa eläviin henkilöihin sekä henkilöiden sukupuoleen, rotuun⁵ tai etniseen alkuperään ja vammaisuuteen, on myös havaittu. Muita huolenaiheita ovat esimerkiksi riittävien ohjeiden ja koulutuksen puute tekoälyä ja ADM-työkaluja käyttäville tapausten käsittelijöille sekä tavat, joilla työnhakijoiden henkilötietoja käytetään profilointi-sovelluksissa. Julkiset työvoimatoimistot voivat käyttää myös muunlaisia välineitä, kuten työvoiman kohdistukseen tarkoitettuja algoritmeja, jotka suosittavat avoimia työpaikkoja työnhakijoille.

Lainvalvonta

Myös lainvalvontaviranomaiset käyttävät säännöllisesti tekoäly- ja ADM-järjestelmiä. Tutkimukset osoittavat kuitenkin, että kasvontunnistusteknologioissa voi esiintyä syrjiviä vinoumia.⁶ Vaikka tekoäly- ja ADM-järjestelmien ei itsessään ilmoitettaisi olevan vinoutuneita, niiden käyttöä arvostellaan siitä, että ne kohdistavat suhteettomasti vähemmistöihin ja vähemmistöjen valvontaan sekä käyttävät etnistä profilointia.⁷ Näyttö tällaisesta käytöstä sisältää usein vain aihetodisteita, ja avoimen tiedon puute siitä, mitä järjestelmiä käytetään ja millä tavoin, estää järjestelmällisen tutkimuksen riskeistä, joita järjestelmät aiheuttavat perusoikeuksille, kuten syrjimättömyydelle. Kuten jäljempänä todetaan (katso kohta 1.4.1.), tekoälyasetuksessa myönnetään poliisille ja lainvalvontaviranomaisille laajoja poikkeuksia, jotka koskevat esimerkiksi kasvokuvia ja muita biometrisia tietoja käsittelevien tunnistusteknologioiden käyttöä.

Sosiaalihuoltosektorit

Useissa eurooppalaisissa sosiaalihuoltojärjestelmissä kokeillaan tai otetaan käyttöön parhaillaan tekoäly- ja ADM-järjestelmiä (esimerkiksi Alankomaissa, Belgiassa, Itävallassa, Portugalissa, Puolassa, Ranskassa, Ruotsissa Sloveniassa ja Tanskassa). Ranskassa perhe-etuksien myöntämisestä vastaavan kansallisen viraston käyttämän järjestelmän tarkoituksena on ennakoida petos- ja virheriskejä edunsaajien keskuudessa tapausten käsittelijöiden tarkastusten kohdennusta varten (La Quadrature du Net, 2023). Kansalaisjärjestöt ovat tuoneet esille syrjintäriskejä, ja esimerkiksi

5 Koska kaikki ihmiset kuuluvat samaan lajiin, Euroopan neuvoston ministerikomitea hylkää teoriat, jotka perustuvat eri "rotujen" olemassaoloon. Tässä julkaisussa termiä "rotu" käytetään kuitenkin sen varmistamiseksi, etteivät henkilöt, joita yleisesti ja virheellisesti rodullistetaan, jää lainsäädännön tarjoaman suojan ja politiikan täytäntöönpanon ulkopuolelle.

6 CSO Liberty arvosteli Etelä-Walesin poliisin käyttämää kasvontunnistussovellusta esimerkiksi sukupuoleen ja/tai rotuun perustuvasta syrjinnästä, koska positiivisten tulosten suhde oli suurempi naisten ja/tai tummaihoisten ja etnisiin vähemmistöihin kuuluvien henkilöiden kasvojen kohdalla. Katso Yhdistyneen kuningaskunnan muutoksenhakutuomioistuimen päätös asiassa *R (Bridges) v. Chief Constable of South Wales Police* ([2020] EWCA Civ 1058), jossa korostettiin, että Etelä-Walesin poliisin "yhdenvertaisuusvaikutusten arviointi oli ilmeisen puutteellinen ja perustui oikeuserhdykseen (koska siinä ei tunnistettu välillisen syrjinnän riskiä)" ja että sen "myöhempi lähestymistapa automaattisen kasvontunnistuksen käyttöön liittyvän välillisen syrjinnän arviointiin on virheellinen".

7 Näin tapahtuu esimerkiksi silloin, kun poliisivoimat säilyttävät järjestelmissään kuvia virheellisistä osumista suhteettoman pitkään tai valvovat suhteettomasti paikkoja, jossa asuu etnisiä taustoja omaavia yhteisöjä, mikä vaikuttaa ennakointityökaluihin ja vahvistaa täten entisestään näiden yhteisöjen joukkovalvontaa.

sukupuoleen, perhetilanteeseen, ikään ja vammaisuuteen liittyvää syrjintää koskeva tapaus on viety Ranskan korkeimman hallinto-oikeuden (Conseil d'État) käsiteltäväksi (katso Conseil d'État, 2024). Alankomaissa samankaltainen järjestelmä johti hallituksen eroon vuonna 2021, kun sen todettiin syrjineen tuensaajia rodun, etnisen alkuperän ja kansalaisuuden perusteella.⁸ Koulutuksen osalta Alankomaiden koulutusvirasto (Dutch Executive Agency of Education) käytti riskiprofilointialgoritmia tukemaan opintotukien saajiin liittyvien petosten havaitsemista (Algorithm Audit, 2024a). Virasto tunnusti, että "[o]piskelijoille, joilla oli muu kuin eurooppalainen maahanmuuttajatausta, annettiin riskiprofilin perusteella korkeammat riskipisteet, ja heidät valittiin muita useammin manuaalisesti kotitarkastuskäynnille", mikä loi välillistä syrjintää (Algorithm Audit, 2024b). Belgiassa OASIS-järjestelmä lakkautettiin vuonna 2023 sen jälkeen, kun tutkijat osoittivat köyhyyteen perustuvan syrjinnän riskin. Tällaisissa järjestelmissä yhdistetään eri viranomaisilta saatuja tietoja (esimerkiksi verotuksesta, työllisyydestä, perhe-etuuksista ja eläkkeistä) ja joskus myös yksityisten yritysten keräämiä tietoja (esimerkiksi energiankulutuksesta), joita käytetään käyttäjien profilointiin ja riskipisteiden määrittämiseen.

Muuttoliike ja kansalaisuus

Jotkut eurooppalaiset julkishallinnot käyttävät tekoälyä ja ADM-järjestelmiä myös muuttoliikkeeseen liittyvien päätösten tukena. Esimerkkejä tästä ovat kansalaisuutta, turvapaikkaa tai oleskelua koskevat päätökset. Raportoituja käyttötarkoituksia ovat esimerkiksi kielen tunnistaminen ja arviointi, henkilöasiakirjoihin liittyvien petosten havaitseminen, tapausten hallinta, vuorovaikutus maahanmuuttajien kanssa muun muassa chattibottien avulla, muuttoliike-ennusteet ja rajavalvontateknologiat (Euroopan muuttoliikeverkosto ja Taloudellisen yhteistyön ja kehityksen järjestö, 2022; kansallisten ihmisoikeusinstituutioiden eurooppalainen verkosto, 2024; McGregor ja Molnar, 2023). Vaikka näitä teknologioita väitetään käytettävän muuttoliikkeen hallinnan tehostamiseen, ne voivat vaikuttaa kielteisesti maahanmuuttajien oikeuksiin vahvistamalla syrjintää, luomalla yksityisyyteen ja tietosuojaan liittyviä huolenaiheita ja estämällä maahanmuuttajia hakemasta suojelua (kansallisten ihmisoikeusinstituutioiden eurooppalainen verkosto, 2014). YK:n rasismiin, rotusyrjinnän, muukalaisvihan ja niihin liittyvän suvaitsemattomuuden nykymuotoja käsittelevän erityisraportoin mukaan "digitaaliset rajat" käsittävät erilaisia hallintoinfrastruktuureja, jotka "tukeutuvat yhä enemmän koneoppimiseen, massadataan, automaattisiin algoritmisiin päätöksentekojärjestelmiin, ennustavaan analytiikkaan ja niihin liittyviin digitaalisiin teknologioihin" (YK:n erityisraportoiija, 2021). Käytetyt teknologiat sisältävät muun muassa "tunnistusjärjestelmiä, kasvontunnistusjärjestelmiä, maaperään asetettavia antureita, videovalvontadrooneja, biometrisiä tietokantoja ja jopa viisumi- ja turvapaikkapäätösprosesseja" (ibid.).

8. SyRI-järjestelmän katsottiin myös loukanneen kohtuuttomasti loppukäyttäjien oikeutta yksityisyyteen, koska se käsitteli eri valtion virastoilta saatuja henkilötietoja.

Veroviranomaiset

Vaikka veroviranomaisten käyttämistä tekoäly- ja ADM-järjestelmistä on yleisesti ottaen saatavilla vähemmän tietoa, useat Euroopan maat kokeilevat parhaillaan tällaisia teknologioita lähinnä petosten havaitsemista varten. Tekoäly- ja ADM-järjestelmiä käytetään esimerkiksi epäilyttävien tiedostojen merkitsemiseen lisätutkimuksia varten. Viranomaiset saattavat olla haluttomia paljastamaan tietoja näistä järjestelmistä, koska pelätään, että käyttäjät pystyvät ”hyväksikäyttämään” järjestelmää. Syrjinnän riskejä on olemassa, koska petosten havaitsemisjärjestelmät voivat vaikuttaa suhteettomasti tiettyihin yksilöihin kiellettyjen syrjäntäperusteiden perusteella. Tällaisia voivat olla esimerkiksi sosioekonominen asema, maahanmuuttajatausta tai maantieteellinen sijainti, vaikka näitä kategorioita ei olisi nimenomaisesti käytetty riskitekijöinä näissä järjestelmissä. Edellä mainitussa Alankomaiden SyRI-tapauksessa veroviranomaiset epäilivät 26 000 perhettä sosiaaliturvaetuuksiin liittyvistä petoksista. Algoritmi merkitsi epäilyiksi suhteettomasti etnisiä vähemmistöjä ja muita kuin Alankomaiden kansalaisia.

Yksityinen sektori

Yksityisellä sektorilla tekoäly- ja ADM-järjestelmiä käyttävät esimerkiksi henkilöstöpalveluyritykset. Käyttösovelluksia ovat muun muassa työnhakijoiden seulonta ja profilointi, työhakemusten ja profiilien yhteensovittaminen tai avointen työpaikka-ilmoitusten laatiminen ja kääntäminen. Näihin käyttösovelluksiin voi liittyä eriasteisia syrjäntäriskejä. Vastaavasti tekoälyä ja ADM-järjestelmiä käytetään näkyvästi rahoitus- ja kaupan alalla, kuten markkinoinnissa, asiakaspalvelussa (esimerkiksi chattibotit), dynaamisessa hinnoittelussa ja tapahtumien valvonnassa. Esimerkiksi Suomessa rahoituspalveluyritys ei myöntänyt asiakkaalle lainaa verkossa. Päätös perustui lainan luottotappioriskin arvioinnissa käytettyyn pisteytysjärjestelmään, joka otti pisteytyksessä huomioon esimerkiksi sukupuolen, iän, äidinkielen ja asuinpaikan (Algorithm Watch, 2018). Suomen yhdenvertaisuus- ja tasa-arvolautakunta (2017) totesi, että kyseessä oli välitön syrjäntä, joka perustui useisiin kiellettyihin syrjäntäperusteisiin.

1.4. Olemassa olevat oikeudelliset kehykset

Tässä osiossa analysoidaan muun muassa nykyisiä asetuksia, lakeja, oikeuskäytäntöä, menettelyjä, käytäntöjä ja instituutioita ja tarkastellaan, miten niissä käsitellään tekoälyn ja ADM-järjestelmien käytöstä johtuvaa syrjäntää.

1.4.1. Tekoälyn ja ADM-järjestelmien hallinnointi

Euroopan neuvosto

Euroopan neuvoston tasolla on useita välineitä, joilla pyritään käsittelemään algoritmista syrjäntää tai joita voidaan käyttää siihen.

Ensinnäkin Euroopan neuvoston tekoälyä koskevalla puiteyleissopimuksella, joka hyväksyttiin vuonna 2024 mutta jota ei ole vielä ratifioitu, ”pyritään varmistamaan, että tekoälyjärjestelmien elinkaareen kuuluvat toimet ovat täysin yhdenmukaisia

ihmisoikeuksien, demokratian ja oikeusvaltioperiaatteen kanssa”. Tekoälyä koskevan puiteyleissopimuksen johdanto-osassa tunnustetaan, että tekoälyllä voidaan ”edistää... sukupuolten tasa-arvoa ja naisten ja tyttöjen oikeuksia ja mahdollisuuksia”. Samalla siinä kuitenkin ilmaistaan, että sopimuspuolet ”ovat huolissaan syrjinnän riskeistä digitaalisissa yhteyksissä, erityisesti tekoälyjärjestelmiin liittyvissä yhteyksissä, ja niiden mahdollisista vaikutuksista eriarvoisuuden syntymiseen tai pahenemiseen, mukaan lukien naisten ja haavoittuvassa asemassa olevien yksilöiden kokemaa eriarvoisuutta, joka liittyy heidän ihmisoikeuksiensa toteutumiseen ja heidän täysipainoiseen, yhdenvertaiseen ja tosiasialliseen osallistumiseensa taloudellisiin, sosiaalisiin, kulttuurisiin ja poliittisiin asioihin”.

Tekoälyä koskevan puiteyleissopimuksen tasa-arvoa ja syrjimättömyyttä koskevassa 10 artiklassa todetaan, että ”kukin sopimuspuoli ottaa käyttöön tai pitää voimassa toimenpiteet, joilla pyritään varmistamaan, että tekoälyjärjestelmien elinkaareen⁹ kuuluvissa toimissa kunnioitetaan yhdenvertaisuutta, myös sukupuolten tasa-arvoa, ja syrjintäkieltoa, siten kuin niistä säädetään sovellettavassa kansainvälisessä ja kansallisessa lainsäädännössä”. Artiklassa todetaan myös, että ”kukin sopimuspuoli sitoutuu ottamaan käyttöön tai pitämään voimassa toimenpiteet, joilla pyritään poistamaan eriarvoisuutta ja saavuttamaan oikeudenmukaisia ja tasapuolisia tuloksia tekoälyjärjestelmien elinkaareen kuuluvissa toimissa, sovellettavien kansallisten ja kansainvälisten ihmisoikeusvelvoitteidensa mukaisesti”. Nämä velvoitteet edellyttävät, että jäsenvaltiot tarkistavat ja tarvittaessa uudistavat lainsäädäntöään varmistaakseen, että yhdenvertaisuuslainsäädäntö kattaa algoritmisen syrjinnän. Kuten perusteluosan kohdassa 77 on kuvattu, puiteyleissopimuksessa asetetaan valtioille myös positiivinen velvoite toteuttaa toimenpiteitä rakenteellisen ja historiallisen eriarvoisuuden poistamiseksi tekoälyjärjestelmien elinkaaren aikana. Perusteluosan kohdissa 72–73 vahvistetaan, että puiteyleissopimukseen sisältyvä algoritmisen syrjinnän kieltä perustuu algoritmisen syrjinnän ihmisoikeuspohjaiseen lähestymistapaan, jossa yhdistyvät kansainväliset ja alueelliset ihmisoikeuskehykset. Perusteluosan kohdissa 75–76 on luettelo tunnetuista algoritmisten vinoumien lähteistä ja tyypeistä. Perusteluosassa mainitaan, että ”tekoälyjärjestelmiin voidaan tahallisesti tai tahattomasti sisällyttää vinoumia eri tavoin niiden elinkaaren eri vaiheissa”, ja näitä tapoja ovat seuraavat:

- ▶ kehittäminen (”kehittäjien tietoisten tai tiedostamattomien stereotyyppien tai vinoumien vuoksi”), mallintaminen (”mahdolliset vinoumat, jotka sisältyvät järjestelmien rakentamisessa käytettyyn malliin”)
- ▶ data (epätarkat tai tietoja riittämättömästi edustavat datajoukot koulutus-, aggregaatio- tai arviointivaiheessa)
- ▶ käyttöönotto (”vinoumat, joita syntyy, kun tällaisia järjestelmiä käytetään tosimaailman olosuhteissa”)

9. Katso puiteyleissopimuksen selitysmuistion kohdat 14–15: ”Tämä viittaus elinkaareen takaa kattavan lähestymistavan tekoälyyn liittyvien riskien ja ihmisoikeuksiin, demokratiaan ja oikeusvaltioperiaatteeseen kohdistuvien haittavaikutusten käsittelyyn, sillä se kattaa kaikki tekoälyjärjestelmiin liittyvät toiminnan vaiheet.” Esimerkkejä sovellettavista toimista ovat seuraavat: ”(1) suunnittelu ja kehitys, (2) tiedonkeruu ja -käsittely, (3) tekoälyjärjestelmien kehittäminen, mukaan lukien mallien rakentaminen ja/tai olemassa olevien mallien hienosäätö tiettyjä tehtäviä varten, (4) testaus, vahvistus ja validointi, (5) toimitus / asettaminen saataville käyttöä varten, (6) käyttöönotto, (7) käyttö ja valvonta sekä (8) käytöstä poistaminen”.

- ▶ tulkinta ("automaatio tai vahvistusharhat, joissa ihmiset saattavat luottaa perusteettomasti koneisiin ja teknologisiin laitteisiin tai valita omia näkemyksiään tukevia tietoja.. huomioimatta omia, mahdollisesti ristiriitaisia arvioitaan ja validoimalla algoritmien tuotokset kyseenalaistamatta niitä")
- ▶ tekninen vinouma ("joka syntyy, kun koneoppimisen käytön ongelmat johtavat lisävinoumiin, joita ei esiinny järjestelmän koulutuksessa tai päätöksenteossa käytetyssä datassa")
- ▶ sosiaalinen vinouma ("tekoälyjärjestelmien elinkaaren aikana tapahtuvassa toiminnassa ei oteta asianmukaisesti huomioon yhteiskunnan historiallista tai nykyistä eriarvoisuutta").

Puiteyleissopimukseen sisältyvien velvoitteiden täytäntöönpanon helpottamiseksi Euroopan neuvosto on laatinut "menetelmä[n], jolla varmistetaan yhdenmukainen lähestymistapa [tekoälyjärjestelmien] riskien tunnistamiseen, analysointiin ja arviointiin sekä vaikutusten arviointiin ihmisoikeuksien toteutumisen, demokratian toiminnan ja oikeusvaltioperiaatteiden noudattamisen kannalta". Tähän menetelmään viitataan nimellä HUDERIA. (Euroopan neuvosto, 2022a). Se koostuu kontekstiin perustuvasta riskianalyysistä, sidosryhmien osallistamisprosessista, riskien ja vaikutusten arvioinnista sekä lieventämissuunnitelmasta, ja se edellyttää iteratiivista tarkistusta. Vaikka lähestymistapa eroaa tekoälyasetuksen mukaisesta riskiperusteisesta luokittelusta (katso jäljempää Euroopan unionia koskeva osio), HUDERIA-menetelmässä käytetään samankaltaisia elementtejä "riskien ja vaikutusten tunnistuksen, arvioinnin, ehkäisyn ja lieventämisen porrastetulla ja eriytettyllä lähestymistavalla suhteessa siihen, kuinka vakavia ja todennäköisiä ihmisoikeuksiin, demokratiaan ja oikeusvaltioperiaatteen kohdistuvat riskit ja vaikutukset ovat" (Euroopan neuvosto, 2024b). Esimerkiksi riskitasojen määrittämisessä on otettava huomioon mahdollisten haittojen laajuus, vakavuus, peruutettavuus ja todennäköisyys. Myös "nollakysymyksiä", jotka koskevat tekoäly- ja ADM-järjestelmien soveltuvuutta ja riittävyyttä tiettyjen tehtävien suorittamiseen, on tarkasteltava. Tämän tarkastelun mahdollisena seurauksena on, että tällaisten järjestelmien kehittämisestä tai käyttöönotosta luovutaan, jos ihmisoikeusloukkausten riskit ovat suuremmat kuin mahdolliset hyödyt. Riskien ja vaikutusten HUDERIA-arviointimenetelmällä pyritään varmistamaan saumaton yhteensopivuus "alan nykyisten vaatimustenmukaisuuskäytäntöjen kanssa" (Euroopan neuvosto, 2022a), ja se voi edistää julkisten ja yksityisten toimijoiden algoritmisen syrjinnän ehkäisemistä. Vaikka puiteyleissopimuksen 16 artiklassa määrätään riskin- ja vaikutustenhallintajärjestelmän käyttöönotosta, HUDERIAN noudattaminen ei ole pakollista puiteyleissopimuksen velvoitteiden täyttämiseksi. Vaikka HUDERIA ei ole oikeudellisesti sitova eikä tarjoa puiteyleissopimuksen tulkintaohjeita, se on silti hyödyllinen väline algoritmiseen syrjintään liittyvien riskien tunnistamisessa, arvioinnissa ja lieventämisessä.

Lisäksi on julkaistu erilaisia suosituksia, joissa sopimuspuolia vaaditaan puuttumaan algoritmiseen syrjintään lainsäädännön ja menettelyjen puitteissa. Esimerkiksi parlamentaarinen yleiskokous kehotti tekoälyn käytöstä johtuvan syrjinnän ehkäisyä koskevassa päätöslauselmassaan 2343 (2020) "Preventing discrimination caused by the use of artificial intelligence" jäsenvaltioita:

tarkistamaan syrjintää torjuvaa lainsäädäntöään ja muuttamaan sitä tarvittaessa varmistakseen, että se kattaa kaikki tapaukset, joissa tekoälyn käyttö voi aiheuttaa välitöntä tai välillistä syrjintää, mukaan luettuna läheissyryntää, ja että oikeussuoja on saatavilla kanteluiden tekijöille; oikeussuojassa on kiinnitettävä erityistä huomiota syyttömyysoletaman takaamiseen ja sen varmistamiseen, että syrjinnän uhreille ei koidu kohtuutonta todistustaakkaa (Euroopan neuvosto, parlamentaarinen yleiskokous, 2020a).

Lisäksi päätöslauselmassa kehoitettiin jäsenvaltioita "varmistamaan, että yhdenvertaisuuselimillä on täydet valtuudet käsitellä tekoälyn käytöstä johtuvia yhdenvertaisuuteen ja syrjimättömyyteen liittyviä kysymyksiä". Siinä vaadittiin myös, että hallituksia pyydetään "ilmoittamaan parlamentille ennen [tekoäly- ja ADM-järjestelmien] teknologian käyttöönottoa" ja että "tällaisen teknologian viranomaiskäyttö... kirjataan järjestelmällisesti julkiseen rekisteriin". On huomattava, että monet näistä yksityiskohtaisista suosituksista eivät sisälly äskettäin hyväksyttyyn tekoälyä koskevaan puiteyleissopimukseen. Syrjinnän torjuntaa, monimuotoisuutta ja osallisuutta käsittelevän ohjauskomitean ja sukupuolten tasa-arvosta vastaavan komitean meneillään oleva työ ministerikomitean tekoälyyn liittyvän yhdenvertaisuuden suosituksesta voi tarjota ratkaisuja näihin puutteisiin (Euroopan neuvosto, tulossa).

Myös muita Euroopan neuvoston instituutioiden julkaisemia välineitä voidaan käyttää algoritmisen syrjinnän käsittelyyn. Esimerkiksi naisiin kohdistuvan väkivallan ja perheväkivallan torjunnan asiantuntijaryhmän (GREVIO) hyväksymässä ensimmäisessä yleisessä suosituksessa naisiin kohdistuvan väkivallan digitaalisesta ulottuvuudesta puututaan erityisesti verkkoväkivallan syrjiiviin muotoihin, kuten verkossa tapahtuvaan seksuaaliseen häirintään, verkossa ja teknologian avulla tapahtuvaan ahdisteluun sekä psykologisen väkivallan digitaalisiin ulottuvuuksiin (Euroopan neuvosto, 2021a).

Mainitsemisen arvoisia ovat myös seuraavat monialaisemmat suositukset tai vastaaviin aiheisiin keskittyvät välineet:

- ▶ teknologian yhdentymistä, tekoälyä ja ihmisoikeuksia käsittelevä Euroopan neuvoston parlamentaarisen yleiskokouksen suositus 2102 (2017) "Technological convergence, artificial intelligence and human rights" (Euroopan neuvosto, parlamentaarinen yleiskokous, 2017)
- ▶ digitaalisen seurannan ja muiden valvontateknologioiden aiheuttamia perusoikeusriskejä koskeva ministerikomitean julistus (Declaration of the Committee of Ministers on Risks to Fundamental Rights stemming from Digital Tracking and other Surveillance Technologies) (Euroopan neuvosto, 2013)
- ▶ tekoälyä ja ihmisoikeuksien suojelua koskeva ihmisoikeusvaltuutetun suositus "Unboxing Artificial Intelligence: 10 steps to protect Human Rights" (Euroopan neuvosto, ihmisoikeusvaltuutettu, 2019)
- ▶ algoritmisten prosessien manipuloivia mahdollisuuksia koskeva ministerikomitean julistus (Declaration of the Committee of Ministers on the manipulative capabilities of algorithmic processes) (Euroopan neuvosto, 2019a)
- ▶ ministerikomitean suositus CM/Rec(2020)1 jäsenvaltioille algoritmisten järjestelmien ihmisoikeusvaikutuksista (Euroopan neuvosto, 2020)
- ▶ ministerikomitean suositus CM/Rec(2022)13 jäsenvaltioille digitaalisen teknologian vaikutuksista sananvapauteen (Euroopan neuvosto, 2022b)

- mediaa ja tietoyhteiskuntaa käsittelevän ohjauskomitean hyväksymä sisällön moderointia koskeva ohjeasiakirja "Content moderation. Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation" (Euroopan neuvosto, 2021b)
- ihmisoikeusvaltuutetun ihmisoikeuksien suojelua koskeva seurantasuositus "Human rights by design: future-proofing human rights protection in the era of AI" (Euroopan neuvosto, ihmisoikeusvaltuutettu, 2023).

Nämä eri välineet tarjoavat tärkeitä ohjeita ihmisoikeuksien ja erityisesti syrjimättömyyden toteutumisen varmistamiseen tekoälyä ja ADM-järjestelmiä käytettäessä. Ne eivät kuitenkaan ole sitovia, ja niiden rajallinen täytäntöönpanokelpoisuus rajoittaa niiden tehokkuutta käytännössä.

Euroopan unioni

EU:n tasolla algoritmista syrjintää käsitellään useissa tekoälyasetuksen määräyksissä (Euroopan unioni, 2024c).¹⁰ Tekoälyasetuksen tarkoituksena on täydentää nykyisiä syrjinnän torjuntaa ja tietosuojaa koskevia puitteita. Siinä käytetään riskiperusteista lähestymistapaa ja mainitaan useita kiellettyjä käytäntöjä ja suuririskisiä järjestelmiä, joihin sovelletaan erityisvaatimuksia. Tämä perustuu tunnustukseen, että "[t]ekoälyn monien hyödyllisten käyttötarkoitusten lisäksi sitä voidaan myös käyttää väärin, ja se voi tarjota uusia ja tehokkaita välineitä käytäntöihin, joihin liittyy manipulointia, hyväksikäyttöä ja sosiaalista valvontaa. Tällaiset käytännöt ovat erityisen haitallisia ja loukkaavia, ja ne olisi kiellettävä, koska ne ovat... unionin arvojen... vastaisia, mukaan lukien oikeus syrjimättömyyteen" (johdanto-osan kappale 28).

Muita järjestelmiä pidetään suuririskisinä, koska ne "voivat loukata oikeutta ihmisarvoon ja syrjimättömyyteen sekä tasa-arvon ja oikeussuojan arvoja" (johdanto-osan kappale 31). Kolmatta tekoälyjärjestelmien luokkaa, jonka riskejä pidetään vähäisempinä, ei ole säännelty yhtä tarkasti. Toisin sanoen "[t]ekoälyjärjestelmästä perusoikeuskirjassa suojatuille perusoikeuksille aiheutuvan haitallisen vaikutuksen laajuus on erityisen merkityksellinen, kun tekoälyjärjestelmä luokitellaan suuririskiseksi" (johdanto-osan kappale 48). Asetuksen johdanto-osan kappaleessa 27 muistutetaan kuitenkin, että tekoälyä käsittelevän korkean tason asiantuntijaryhmän määrittelemää seitsemää periaatetta voidaan soveltaa kaikkiin järjestelmiin niiden riskitasosta riippumatta: nämä periaatteet ovat "ihmisen toimijuus ja ihmisen suorittama valvonta, tekninen vakaus ja turvallisuus, yksityisyyden suoja ja datanhallinta, avoimuus, monimuotoisuus, yhdenvertaisuus ja oikeudenmukaisuus sekä yhteiskunnallinen ja ympäristöön liittyvä hyvinvointi ja vastuuvélvollisuus". Tässä kontekstissa "[m]onimuotoisuudella, syrjimättömyydellä ja oikeudenmukaisuudella tarkoitetaan, että tekoälyjärjestelmiä kehitetään ja käytetään tavalla, jossa on mukana monimuotoisia toimijoita ja jolla edistetään yhtäläisiä mahdollisuuksia, sukupuolten tasa-arvoa ja kulttuurien monimuotoisuutta ja vältetään unionin tai kansallisessa lainsäädännössä kiellettyjä syrjiviä vaikutuksia ja epäoikeudenmukaisia vääristymiä" (ibid.). Tätä periaatetta on tarkoitus soveltaa kaikkiin EU:n markkinoille saatettaviin tekoälyjärjestelmiin niiden riskiluokituksesta riippumatta.

10. Katso myös perusteluosa ja siihen liittyvä tekoälysopimus, joka on saatavilla osoitteessa <https://digital-strategy.ec.europa.eu/en/policies/ai-pact>, viitattu 10. marraskuuta 2025.

Kielletyt tekoälyjärjestelmät

Luettelo kielletyistä järjestelmistä, jotka ovat erityisen tärkeitä syrjinnän kannalta, sisältyy tekoälyasetuksen 5 artiklaan. Asetuksen 5 artiklan a alakohdassa kielletään järjestelmät, joissa käytetään ”subliminaalisia tekniikoita... tai tarkoituksellisesti manipuloivia tai harhaanjohtavia tekniikoita, joiden tavoitteena on vääristää tai jotka olennaisesti vääristävät henkilön tai henkilöryhmän käyttäytymistä heikentäen tuntuvasti tämän kykyä tehdä tietoinen päätös ja saaden tämän tekemään päätöksen, jota tämä ei olisi muussa tapauksessa tehnyt, tavalla, joka aiheuttaa tai kohtuullisella todennäköisyydellä aiheuttaa kyseiselle henkilölle, toiselle henkilölle tai henkilöryhmälle merkittävää haittaa”. Asetuksen 5 artiklan b alakohdassa kielletään tekoäly- ja ADM-järjestelmät, joissa hyödynnetään ”luonnollisen henkilön tai tietyn henkilöryhmän haavoittuvuuksia, jotka liittyvät heidän ikäänsä, vammaansa tai erityiseen sosiaaliseen tai taloudelliseen tilanteeseensa” siten, että henkilöiden käyttäytyminen vääristyy ja aiheuttaa heille haittaa.

Asetuksen 5 artiklan c alakohdassa kielletään sosiaaliset pisteytysjärjestelmät, jotka arvioivat tai luokittelevat henkilöitä tai ryhmiä ”heidän sosiaalisen käyttäytymisensä tai tunnettujen, pääteltyjen tai ennakoitujen henkilökohtaisten ominaisuuksiensa tai luonteenpiirteidensä perusteella” siten, että heitä kohdellaan epäedullisesti sosiaalisissa yhteyksissä, jotka eivät liity asiayhteyteen, tai kohteluun, joka on perusteetonta tai suhteetonta. Asetuksen 5 artiklan d alakohdassa kielletään ”tekoälyjärjestelmän... käyttö luonnollisia henkilöitä koskevien riskinarviointien tekemiseen, jotta voidaan arvioida tai ennustaa sen riskiä, että luonnollinen henkilö syyllistyy rikokseen, kun tämä perustuu yksinomaan luonnollisen henkilön profilointiin tai hänen persoonallisuuspiirteidensä ja persoonallisuusominaisuuksiensa arviointiin”. Asetuksen 5 artiklan e alakohdassa kielletään kasvontunnistusjärjestelmät, jotka luodaan haravoimalla kasvokuvia internetistä tai valvontakamerakuvista.

Asetuksen 5 artiklan f alakohdassa kielletään tekoälyjärjestelmät henkilön tunteiden päättelemiseksi työpaikalla ja oppilaitoksissa. Asetuksen 5 artiklan g ja h alakohdan mukaan ”sellaisten biometristen luokittelujärjestelmien käyttö, joissa luonnolliset henkilöt luokitellaan heidän biometristen tietojensa perusteella rodun, poliittisten mielipiteiden, ammattiliiton jäsenyyden, uskonnollisen tai filosofisen vakaumuksen, seksuaalisen käyttäytymisen tai seksuaalisen suuntautumisen perusteella” on kielletty, kuten myös ”reaaliaikaisten biometristen etätunnistusjärjestelmien käyttö julkisissa tiloissa lainvalvontatarkoituksessa”, paitsi jos ja siltä osin kuin se on ehdottoman välttämätöntä.

Näillä kielloilla puututaan algoritmisen syrjinnän tärkeisiin riskeihin yleiskiellon avulla. Kriitikot ovat kuitenkin huomauttaneet merkittävistä porsaanrei’istä, jotka liittyvät poliiseille ja lainvalvontaviranomaisille myönnettyihin laajoihin poikkeuksiin ja erityisesti 5 artiklan g ja h alakohdassa kuvattuun reaaliaikaiseen kasvontunnistukseen ja biometriseen valvontaan (European Digital Rights, 2024). Euroopan komissio on antanut kiellettyjä tekoälykäytäntöjä koskevia suuntaviivoja, joissa selvennetään tekoälyasetuksen tiettyjen määräysten soveltamisalaa. Suuntaviivoissa todetaan seuraavaa:

vaikka tekoälyjärjestelmää ei kiellettäisikään tekoälyasetuksessa, sen käyttö voi silti olla kiellettyä tai laitonta unionin muun primaari- tai sekundaarilainsäädännön perusteella

(esimerkiksi siksi, että perusoikeuksia ei kunnioiteta tietyssä tapauksessa, kuten tietosuojalainsäädännön mukaisen henkilötietojen käsittelyn oikeusperustan puuttuminen tai unionin lainsäädännössä kielletty syrjintä) (Euroopan komissio, 2025a, kohta 43).

Suuririskiset tekoälyjärjestelmät

Tekoälyasetuksen liitteessä III luetellaan suuririskiset tekoälyjärjestelmät, joita käytetään biometriikassa, kriittisessä infrastruktuurissa (kriittisen digitaalisen infrastruktuurin, tieliikenteen tai vesi-, kaasu-, lämmitys- tai sähköhuollon hallinnassa ja toiminnassa), yleissivistävässä ja ammatillisessa koulutuksessa, työllistämisessä, henkilöstöhallinnossa ja itsenäisen ammatinharjoittamisen mahdollistamisessa, välttämättömien yksityisten palvelujen ja välttämättömien julkisten palvelujen ja etuuksien saatavuudessa ja käytössä, lainvalvonnassa, muuttoliikkeen hallinnassa, turvapaikka-asioissa ja rajavalvonnassa sekä oikeudenhoidossa ja demokraattisissa prosesseissa. Näihin suuririskisiin järjestelmiin sovelletaan jäljempänä lueteltuja erityisiä oikeudellisia vaatimuksia. Asetuksen 6 artiklan 3 kohdassa todetaan, että liitteessä III luetellulla alalla käytettyä tekoälyjärjestelmää ”ei ole pidettävä suuririskisenä, jos se ei aiheuta merkittävää vahingon riskiä luonnollisten henkilöiden terveydelle, turvallisuudelle tai perusoikeuksille, mukaan lukien se, että se ei vaikuta olennaisesti päätöksenteon tulokseen”. Näihin kuuluvat järjestelmät, joiden tarkoituksena on ”suorittaa suppea menettelyllinen tehtävä” tai ”valmistelutehtävä”, ”parantaa aiemmin suoritetun ihmisen toiminnan tulosta” tai ”havaita päätöksentekotapoja tai poikkeamia aiemmista päätöksentekotavoista, eikä sen tarkoituksena ole korvata aiemmin tehtyä ihmisen tekemää arviota tai vaikuttaa siihen ilman asianmukaista ihmisen suorittamaa arviota”. Tämä säännös saattaa jättää 6 artiklan soveltamisalan ulkopuolelle suuren joukon suuririskisillä alueilla käytettäviä tekoälyjärjestelmiä. Lisäksi on olemassa mahdollisuus, että säännöstä käytetään väärin, jotta voidaan välttää noudattamasta tekoälyasetuksessa suuririskisille järjestelmille asetettuja vaatimuksia (katso kohta 2.5.). Tällaisiin järjestelmiin sovelletaan kuitenkin edelleen tietyt velvoitteita: tarjoajien, jotka pitävät tekoälyjärjestelmää muuna kuin suuririskisenä, on kuitenkin laadittava arviointia koskevat asiakirjat ennen järjestelmän markkinoille saattamista tai käyttöönottoa ja toimitettava nämä asiakirjat pyynnöstä kansallisille toimivaltaisille viranomaisille. Niiden on myös rekisteröitävä tällainen tekoälyjärjestelmä tekoälyasetuksen 71 artiklassa kuvattuun EU:n tietokantaan.

Yleiskäyttöiset tekoälymallit

Yleiskäyttöisiä tekoälymalleja (GPAI), joiden määritelmä on 3 artiklan 63 kohdan mukaan ”[tekoälymalli], joka on hyvin yleisluonteinen ja pystyy suorittamaan pätevästi monenlaisia erillisiä tehtäviä riippumatta siitä, miten malli saatetaan markkinoille, ja joka voidaan integroida erilaisiin ketjun loppupään järjestelmiin tai sovelluksiin”, säännellään erityisesti tekoälyasetuksen 53 artiklassa. Tarjoajien on laadittava ja toimitettava pyynnöstä viranomaisille tekniset asiakirjat ja käyttöohjeet, noudatettava EU:n tekijänoikeuslainsäädäntöä ja julkaistava yhteenveto koulutustarkoituksiin käytetyistä tiedoista. Tietyn suorituskyvyn ja laajuuden omaavat GPAI-mallit luokitellaan tekoälyasetuksen 51 artiklan mukaan malleiksi, jotka voivat aiheuttaa ”systeemisiä riskejä”, ja niihin sovelletaan tekoälyasetuksen 55 artiklan mukaisia lisävaatimuksia. Tällaisten järjestelmien tarjoajien on esimerkiksi arvioitava malleja ja suoritettava

adversariaalisia testauksia sekä seurattava ja raportoitava vakavia vaaratilanteita. Jos GPAI on integroitu tekoälyjärjestelmään, kyseiseen järjestelmään sovelletaan myös sen riskiluokkaan (riski, jota ei voida hyväksyä, suuri riski, vähäinen riski tai minimaalinen riski) sovellettavia oikeudellisia vaatimuksia.

Vähäisten riskien tekoälyjärjestelmät

Toisin kuin aiemmin kuvattuihin järjestelmiin, tekoälyjärjestelmiin, joihin liittyy vähäisempiä riskejä, sovelletaan vain avoimuusvaatimuksia. Esimerkiksi chattibottien käyttöönottajien on ilmoitettava käyttäjille, että he ovat vuorovaikutuksessa tekoälyjärjestelmän kanssa, ja tekstiä, kuvia, videoita tai ääntä tuottavien GPAI-mallien käyttöönottajien on merkittävä tuotos keinotekoisesti tuotetuksi tekoälyasetuksen 50 artiklan mukaisesti. Muihin tekoälysovelluksiin ei katsota liittyvän riskejä, tai niiden riskit katsotaan minimaalisiksi, eikä niitä säännellä.

Vaatimustenmukaisuuden aikataulu

Tekoälyasetus on nimensä mukaisesti asetus, joten jäsenvaltioiden ei tarvitse saattaa sitä osaksi kansallista lainsäädäntöään, koska sen määräyksiä sovelletaan suoraan. Asetuksen lopullista voimaantuloa edeltää kuitenkin useita eri virstanpylväitä. Asetuksen 5 artiklan mukaisten käytäntöjen kiellot tulivat voimaan helmikuussa 2025, ja komissio on antanut ohjeita näistä kielloista (Euroopan komissio, 2025a). Vuonna 2025 Euroopan tekoälytoimisto julkaisi käytännössä GPAl-malleille, mukaan lukien systeemisistä riskejä aiheuttaville malleille (Euroopan komissio, 2025b). Sen odotetaan myös laativan ohjeita tekoälyjärjestelmien luokitteluun suuririskisiksi järjestelmiksi sekä mallipohjan suuririskisten tekoälyjärjestelmien perusoikeusvaikutusten arviointia varten (katso jäljempää perusoikeusvaikutusten arviointia koskeva osio). Jäsenvaltioiden oli elokuuhun 2025 mennessä nimettävä kansalliset toimivaltaiset viranomaiset (vähintään yksi ilmoittamisesta vastaava viranomainen ja yksi markkinavalvontaviranomainen), joiden tehtävänä on valvoa tekoälyasetuksen kansallista täytäntöönpanoa. Vuonna 2026 Euroopan tekoälytoimiston on päätettävä EU:n standardointielinten, Euroopan standardointikomitean (CEN) ja Euroopan sähkötekniikan standardointikomitean (CENELEC) laatimista teknisistä standardeista, jotka koskevat riskinhallintajärjestelmiä ja vinoumien ehkäisyä suuririskisissä tekoälyjärjestelmissä (katso jäljempää CEN-CENELEC-osio). Lopuksi suuririskisiä järjestelmiä koskevat säännöt tulevat voimaan elokuussa 2026 tai elokuussa 2027 sellaisten tekoälyjärjestelmien osalta, joita käytetään tuotteina tai tuotteiden turvakomponentteina, jos tuotteelle on tehtävä kolmannen osapuolen suorittama vaatimustenmukaisuudenarviointi tekoälyasetuksen liitteen I unionin yhdenmukaistamislainsäädännön mukaisesti. Jos suuririskiset tekoälyjärjestelmät on jo saatettu markkinoille tai otettu käyttöön elokuuhun 2026 mennessä ja ”kyseisten järjestelmien rakenteessa tapahtuu merkittäviä muutoksia” tai ne on tarkoitettu viranomaisten käytettäväksi, tarjoajien ja käyttöönottajien on toteutettava tarvittavat toimenpiteet tekoälyasetuksen velvollisuuksien noudattamiseksi elokuuhun 2030 mennessä (111 artiklan 2 kohta). Kaikki nämä toimet muokkaavat jäsenvaltioiden valmiuksia ehkäistä ja korjata algoritmista syrjintää.

Datanhallinta

Tietyt tekoälyasetuksen määräykset ovat suoraan merkityksellisiä algoritmisen syrjinnän torjunnan kannalta, koska niissä asetetaan erityisvaatimuksia suuririskisille järjestelmille. ”Dataa ja datanhallintaa” koskevassa 10 artiklassa kuvataan laatukriteerit liitteessä III lueteltujen suuririskisten järjestelmien koulutusta, validointia ja testausta varten. Asetuksen 10 artiklan 2 kohdan f ja g alakohdassa todetaan erityisesti, että tekoäly- ja ADM-järjestelmien tarjoajien on suoritettava ”sellaisten mahdollisten vinoutumien selvittäminen, jotka todennäköisesti vaikuttavat ihmisten terveyteen ja turvallisuuteen, vaikuttavat kielteisesti perusoikeuksiin tai johtavat unionin oikeudessa kiellettyyn syrjintään, erityisesti silloin, kun tuotokset vaikuttavat tulevien toimintojen syöttötietoihin” ja ”asianmukaiset toimenpiteet f alakohdan mukaisesti havaittujen mahdollisten vinoutumien havaitsemiseksi, ehkäisemiseksi ja lieventämiseksi”. Näitä tarkoituksia varten voidaan käsitellä arkaluonteisia henkilötietoja, joiden käytön yleinen tietosuoja-asetus ja tekoälyasetus yleisesti kieltävät, 10 artiklan 5 kohdassa säädetyn poikkeuksen perusteella:

Siinä määrin kuin se on ehdottoman välttämätöntä suuririskisten tekoälyjärjestelmien vinoutumien havaitsemisen ja korjaamisen varmistamiseksi [ja tietyin ehdoin¹¹], tällaisten järjestelmien tarjoajat voivat poikkeuksellisesti käsitellä erityisiä henkilötietoryhmiä edellyttäen, että luonnollisten henkilöiden perusoikeudet ja -vapaudet turvataan asianmukaisilla suojatoimilla.

Tämä poikkeus voi olla hyödyllinen algoritmisen syrjinnän tunnistamisessa, mutta se ei kuitenkaan koske muita kuin suuririskisiä järjestelmiä.

Riskinhallintajärjestelmä

Tekoälyasetuksen 9 artikla edellyttää, että suuririskisten tekoälyjärjestelmien tarjoajat ottavat käyttöön riskinhallintajärjestelmän. Siinä todetaan, että ”[s]uuririskisille tekoälyjärjestelmille on perustettava, pantava täytäntöön ja dokumentoitava riskinhallintajärjestelmä ja sitä on pidettävä yllä” ja tämä riskinhallintajärjestelmä on käsitettävä ”iteratiiviseksi prosessiksi, joka jatkuu suunnitellusti suuririskisen tekoälyjärjestelmän koko elinkaaren ajan”. Tarjoajia pyydetään erityisesti toteuttamaan ”niiden tunnettujen ja kohtuudella ennakoitavissa olevien riskien tunnistaminen ja analysointi, joita suuririskinen tekoälyjärjestelmä voi aiheuttaa liittyen

11. Seuraavia ehtoja sovelletaan: ”a) vinoutumia ei voida tehokkaasti havaita ja korjata käsittelemällä muita tietoja, kuten synteettisiä tai anonymisoituja tietoja; b) erityisiin henkilötietoryhmiin sovelletaan henkilötietojen uudelleenkäyttöä koskevia teknisiä rajoituksia ja viimeisintä kehitystä edustavia tietoturvaa ja yksityisyyden suojaa parantavia toimenpiteitä, kuten pseudonymisointia; c) erityisiin henkilötietoryhmiin kohdistetaan toimenpiteitä, joilla varmistetaan, että käsitellyt henkilötiedot turvataan ja suojataan asianmukaisilla suojatoimilla noudattaen, mukaan lukien tietoihin pääsyn tiukka valvonta ja dokumentointi, jotta vältetään väärinkäyttö ja varmistetaan, että ainoastaan valtuutetuilla henkilöillä on pääsy kyseisiin henkilötietoihin asianmukaista salassapitovelvollisuutta soveltaen; d) erityisiä henkilötietoryhmiä ei välitetä, siirretä tai muutoin saateta muiden osapuolten saataville; e) erityisiä henkilötietoryhmiä poistetaan, kun vinoutuma on korjattu tai kun henkilötietojen säilytysaika on päättynyt, sen mukaan kumpi ajankohdista on aikaisempi; f) asetusten (EU) 2016/679 ja (EU) 2018/1725 sekä direktiivin (EU) 2016/680 mukaisiin käsittelytoimia koskeviin selosteisiin sisältyvät syyt, joiden vuoksi erityisten henkilötietoryhmien käsittely oli ehdottoman välttämätöntä vinoutumien havaitsemiseksi ja korjaamiseksi ja miksi tätä tavoitetta ei voitu saavuttaa käsittelemällä muita tietoja.”

terveyteen, turvallisuuteen tai perusoikeuksiin, kun suuririskistä tekoälyjärjestelmää käytetään käyttötarkoituksensa mukaisesti”. Niiden on myös suoritettava ”niiden riskien arviointi, joita voi syntyä, kun suuririskistä tekoälyjärjestelmää käytetään käyttötarkoituksensa mukaisesti ja kohtuudella ennakoitavissa olevissa väärinkäyttöolosuhteissa” sekä ”muiden mahdollisesti syntyvien riskien arviointi 72 artiklassa tarkoitettusta markkinoille saattamisen jälkeisestä seurantajärjestelmästä kerättyjen tietojen analysoinnin perusteella”. Lisäksi niiden velvoitteisiin kuuluu ”sellaisten asianmukaisten ja kohdennettujen riskinhallintatoimenpiteiden hyväksyminen, joiden tarkoituksena on puuttua... tunnistettuihin riskeihin”. Nämä velvoitteet pannaan täytäntöön CEN-CENELECin kehittämien teknisten standardien avulla. Siksi CEN-CENELECin julkaisema riskinhallintastandardi on kriittisen tärkeä väline perusoikeuksiin kohdistuvien riskien hallinnassa, mukaan lukien algoritmisen syrjinnän ehkäisemisessä ja siltä suojautumisessa. Vaikka standardien noudattaminen ei olekaan pakollista, asetuksen 40 artiklan mukainen suuririskisten järjestelmien olettama tiettyjen vaatimusten noudattamisesta on kuitenkin olemassa, ja alan on siksi suositeltavaa noudattaa näitä standardeja. On kuitenkin tärkeää mainita, että tämä olettama tiettyjen vaatimusten noudattamisesta rajoittuu yksinomaan tekoälyasetuksen suuririskisten tekoälyjärjestelmien vaatimuksiin, eikä se kata EU:n perusoikeuslainsäädäntöä tai syrjinnän vastaista lainsäädäntöä.

Muut vaatimukset

Näitä määräyksiä täydennetään useilla suuririskisten tekoälyjärjestelmien tarjoajia koskevilla velvoitteilla, kuten teknistä dokumentaatiota koskevalla 11 artiklalla, tietojen säilyttämistä koskevalla 12 artiklalla, avoimuutta koskevalla 13 artiklalla, ihmisen suorittamaa valvontaa koskevalla 14 artiklalla, tarkkuutta, vakautta ja kyberturvallisuutta koskevalla 15 artiklalla sekä laadunhallintajärjestelmiä koskevalla 17 artiklalla. Esimerkiksi 15 artiklan 4 kohta edellyttää, että suuririskisten järjestelmien, ”jotka jatkavat oppimista markkinoille saattamisen tai käyttöönoton jälkeen”, tarjoajien ”on kehitettävä [järjestelmiä] siten, että poistetaan tai vähennetään mahdollisimman suurelta osin riski mahdollisesti vinoutuneista tuotoksista, jotka vaikuttavat tulevien toimintojen syöttötietoihin (palautesilmukat) ja varmistetaan, että tällaisiin palautesilmukoihin puututaan asianmukaisin lieventävin toimenpitein”. Nämä velvoitteet voivat helpottaa nykyisten syrjinnän vastaisten säännösten täytäntöönpanoa.

EU:n tietokanta

Asetuksen 49 artiklan 1 ja 2 kohdassa tarjoajat veloitetaan rekisteröimään tekoälyjärjestelmät asetuksen 71 artiklassa kuvattuun EU:n tietokantaan. Lisäksi 49 artiklan 3 kohdassa todetaan, että ”käyttöönottajien, jotka ovat viranomaisia, unionin toimielimiä, elimiä, toimistoja tai virastoja tai niiden puolesta toimivia henkilöitä, on rekisteröidyttävä, valittava järjestelmä ja rekisteröitävä sen käyttö”. Jos tekoälyjärjestelmiä käytetään lainvalvonnan, muuttoliikkeen hallinnan, turvapaikka-asioden ja rajavalvonnan aloilla, rekisteröintivelvoitteet olisi täytettävä EU:n tietokannan suojatussa ei-julkisessa osiossa, jota vain Euroopan komissio ja kansalliset viranomaiset voivat käyttää. Asetuksen 71 artiklan mukaan tietokannan muiden osioiden on oltava ”julkisesti saatavilla käyttäjävälisellä tavalla”. Julkinen rekisteri helpottaa tutkimuksia tekoäly- ja ADM-järjestelmien, mukaan lukien julkishallinnon käyttämien

järjestelmien, syrjivistä vaikutuksista sekä mahdollisten uhrien tehtävää saada *prima facie* -näyttöä syrjinnästä. Vaikka EU:n laajuinen tietokanta on hyvä alku, se ei kuitenkaan yksin riitä haastamaan tekoälyn ja ADM-järjestelmien syrjivää käyttöä. Lisäksi kansalaisjärjestöt ovat ilmaisseet huolensa siitä, että rekisteriin sisältyvät lainvalvontaa, muuttoliikkeen hallintaa, turvapaikka-asioita ja rajavalvontaa koskevat tiedot eivät ole julkisesti saatavilla (Access Now, 2024).

Perusoikeusvaikutusten arviointi

Asetuksen 27 artiklan mukaan tiettyjen käyttöönnottajien – lähinnä sellaisten, jotka ovat julkisoikeudellisia laitoksia tai julkisia palveluja tarjoavia yksityisiä yhteisöjä, mutta myös sellaisten suuririskisten tekoälyjärjestelmien käyttöönnottajien, jotka on tarkoitettu käytettäväksi luonnollisten henkilöiden luottokelpoisuuden arviointiin tai heidän luottopisteytyksensä määrittämiseen tai riskinarviointiin ja hinnoitteluun sairaus- ja henkivakuutusten tapauksessa – on suoritettava perusoikeusvaikutusten arviointi (FRIA) ennen suuririskisten tekoälyjärjestelmien käyttöä. Nämä arvioinnit sisältävät kuvauksen käyttöönnottajan prosesseista ja ajanjaksosta, jonka kuluessa tekoälyjärjestelmää on tarkoitus käyttää, ja kuinka usein, niiden luonnollisten henkilöiden ja ryhmien luokat, joihin tekoälyjärjestelmän käyttö todennäköisesti vaikuttaa kyseisessä kontekstissa, erityiset sellaisen haitan riskit, joka todennäköisesti kohdistuu yksilöityihin luokkiin kuuluviin luonnollisiin henkilöihin tai henkilöryhmiin, kuvauksen ihmisen suorittamien valvontatoimenpiteiden toteuttamisesta sekä toimenpiteet, jotka on toteutettava kyseisten riskien toteutuessa, mukaan lukien sisäistä hallintoa ja valitusmekanismeja koskevat järjestelyt. FRIA toimitetaan markkinaevalvontaviranomaisille, ja tiivistelmä on rekisteröitävä EU:n tietokantaan (liite IV, jakso C, kohta 4). FRIA-arviointien käyttömahdollisuus antaa yksilöille ja yhdenvertaisuuselimille paremman mahdollisuuden haastaa syrjiviä tekoäly- ja ADM-järjestelmiä. Lisäksi, kuten johdanto-osan kappaleessa 96 todetaan,

suuririskisten tekoälyjärjestelmien käyttöönnottajat voisivat ottaa asiaankuuluvat sidosryhmät, mukaan lukien sellaisten henkilöryhmien edustajat, joihin tekoälyjärjestelmä todennäköisesti vaikuttaa, riippumattomat asiantuntijat ja kansalaisjärjestöt, mukaan tällaisen vaikutustenarvioinnin toteuttamiseen sekä riskien toteutuessa toteutettavien toimenpiteiden suunnitteluun, etenkin jos tekoälyjärjestelmiä käytetään julkisella sektorilla.

FRIA-arviointeja täydentää 60 artikla. Siinä kuvataan puitteet, joilla tarjoajat voivat (käyttöönnottajien kanssa tai ilman heitä) testata suuririskisiä tekoälyjärjestelmiä tosimaailman olosuhteissa. Vakavista vaaratilanteista on ilmoitettava kansalliselle markkinaevalvontaviranomaiselle (60 artiklan 7 kohta ja 73 artikla), ja niihin on puututtava välittömästi lieventämistoimenpiteillä. Se, missä määrin asiaankuuluvat sidosryhmät, kuten yhdenvertaisuuselimet, voivat tutustua testaustuloksiin ja lieventämistoimenpiteisiin, on kuitenkin epäselvää ja saattaa riippua kansallisesta kontekstista. Toisaalta 71 artiklan 4 kohdassa todetaan, että ”edellä olevan 60 artiklan mukaisesti rekisteröityjen [testaukseen liittyvien] tietojen on oltava ainoastaan markkinaevalvontaviranomaisten ja komission saatavilla, paitsi jos mahdollinen tarjoaja tai tarjoaja on antanut suostumuksensa näiden tietojen asettamiseen myös yleisön saataville”. Siinä itse asiassa jätetään 60 artiklan 4 kohdan c alakohdan mukaisesti

rekisteröidyt testaustiedot saatavuutta ja julkisuutta koskevien vaatimusten ulkopuolelle. Toisaalta 77 artiklan 1 kohdassa todetaan seuraavaa:

Kansallisilla viranomaisilla tai elimillä, jotka valvovat tai panevat täytäntöön perusoikeuksien, mukaan lukien oikeus syrjimättömyyteen, suojelua koskevan unionin oikeuden mukaisia velvoitteita liitteessä III tarkoitettujen suuririskisten tekoälyjärjestelmien käytön yhteydessä, *on oltava valtuudet pyytää ja saada käyttöönsä tämän asetuksen nojalla luotu tai ylläpidetty dokumentaatio* ymmärrettävällä kielellä ja saavutettavassa muodossa, kun dokumentaation käyttö on tarpeen, jotta ne voisivat suorittaa tehokkaasti tehtävänsä omalla lainkäyttöalueellaan. (Korostus lisätty).

Perusoikeuksia suojelevat viranomaiset (77 artiklan mukaiset elimet)

Tekoälyasetuksen 77 artikla ”Perusoikeuksia suojelevien viranomaisten valtuudet” on erityisen tärkeä, koska siinä kuvataan täytäntöönpanokehykset. Kuten edellä on selitetty, 77 artiklan 1 kohdassa annetaan jäsenvaltioiden nimeämille perusoikeuksien valvontaviranomaisille valtuudet pyytää ja saada käyttöönsä tekoälyasetuksen nojalla luotu tai ylläpidetty dokumentaatio, kun se on tarpeen. Näin ollen siinä annetaan useille perusoikeuksia suojeleville viranomaisille oikeus tutustua tekoälyasetuksen nojalla laadittuihin asiakirjoihin, kuten riskinhallintasuunnitelmiin, vaikutustenarviointeihin tai muihin asiakirjoihin, jotka ovat välttämättömiä viranomaisten tehtävän tehokkaan suorittamisen kannalta. Euroopan komissio on antanut laajan tulkinnan asianomaisten viranomaisten luettelosta. Se ei rajoitu vain elimiin, jotka yleensä ymmärretään ihmisoikeusrakenteisiin kuuluviksi, kuten yhdenvertaisuuselimiin, kansallisiin ihmisoikeusinstituutioihin ja oikeusasiamiehiin, vaan siihen kuuluvat myös tietosuojaviranomaiset, kuluttajansuojaviranomaiset, lastensuojeluviranomaiset, työlaainsäädäntöviranomaiset, mediavalvontaviranomaiset ja vaalien rehellisyyden varmistamisesta vastaavat viranomaiset. Viranomaisten, joille on myönnetty 77 artiklan mukaiset käyttöoikeudet, on joko valvottava tai pantava täytäntöön sovellettavaa unionin lainsäädäntöä (tai tehtävä molempia). Komissio katsoo, että 77 artiklan tarkoituksena ei ole kattaa kansallisia tuomioistuimia tai muita oikeusviranomaisia niiden toimiessa oikeudellisessa ominaisuudessaan, sillä niillä on jo EU:n perusoikeuskirjan 47 artiklan mukaiset käyttöoikeudet. Jäsenvaltioiden on ilmoitettava julkisesti 77 artiklan 1 kohdan mukaisista kansallisista viranomaisista tai elimistä ja ilmoitettava niistä myös Euroopan komissiolle. Kun otetaan huomioon yhdenvertaisuuselimille annetut uudet valtuudet (katso jäljempää yhdenvertaisuuselinten roolia koskeva osio), näillä elimillä voi olla merkittävä rooli tekoälyasetuksen täytäntöönpanossa algoritmisen syrjinnän osalta.

Lisäksi 77 artiklan 3 kohdassa todetaan seuraavaa:

Jos 1 kohdassa tarkoitettu dokumentaatio ei riitä sen selvittämiseen, onko perusoikeuksia suojaavan unionin oikeuden mukaisia velvoitteita rikottu, 1 kohdassa tarkoitettu viranomainen tai elin voi esittää markkinavalvontaviranomaiselle perustellun pyynnön järjestää suuririskisen tekoälyjärjestelmän testaus teknisin keinoin. Markkinavalvontaviranomaisen on järjestettävä testaus pyynnön esittävän viranomaisen tai elimen tiiviillä myötävaihtokutsella kohtuullisen ajan kuluessa pyynnön esittämisestä.

Lisäksi tekoälyasetuksen 73 artiklan 7 kohdan mukaan markkinavalvontaviranomaisen on ilmoitettava tietoonsa tulleesta suuririskiseen tekoälyjärjestelmään liittyvästä vakavasta vaaratilanteesta, mukaan lukien syrjinnästä, asiaankuuluville kansallisille

viranomaisille tai elimille. Asetuksen 79 artiklan 2 kohdan mukaan markkinavalvonta-viranomaisen on ilmoitettava asiasta asiaankuuluville viranomaisille ja tehtävä tiivistä yhteistyötä niiden kanssa, jos se tunnistaa riskialttiiseen tekoälyjärjestelmään liittyvän perusoikeusriskin, mukaan lukien mahdollinen syrjintä. Asiaankuuluvien toimijoiden on myös tehtävä tarvittaessa yhteistyötä viranomaisten kanssa. Asetuksen 82 artiklan 1 kohdan mukaan markkinavalvontaviranomaisen on kuultava perusoikeuksia suojelevia kansallisia viranomaisia, jos se toteaa, että vaatimustenmukainen suuririskinen tekoälyjärjestelmä aiheuttaa riskin perusoikeuksille, kuten syrjinnän riskin, jos kyseinen perusoikeus on merkityksellinen kyseisen viranomaisen tai elimen tehtävän kannalta.

Tekoälyyn liittyvä vastuu

Euroopan komissio ehdotti vuonna 2022 erityistä direktiiviä tekoälyyn liittyvästä vastuusta. Sillä pyrittiin puuttumaan tekoälyyn liittyviin oikeussuojan saatavuutta koskeviin haasteisiin. Ehdotuksena oli, että EU:ssa yhdenmukaistettaisiin tekoälyyn liittyvien vahinkojen vastuusäännöt sopimuksenulkoisia tuottamukseen perustuvia siviilioikeudellisia vahingonkorvausvaateita varten. Direktiiviehdotus kuitenkin peruutettiin vuonna 2025, minkä vuoksi on entistäkin kiireellisempää käsitellä tekoälyyn liittyviä vastuukysymyksiä kansallisella ja Euroopan tasolla sekä muilla keinoilla. Tekoälyyn liittyvää vastuuta koskevassa direktiiviehdotuksessa tunnustettiin ”tekoälyn erityispiirteet, joihin liittyy ongelmia nykyisten vastuusääntöjen kannalta”. Direktiivissä ehdotettiin, että ”kevennetään todistustaakkaa hyvin kohdennetulla ja oikeasuhteisella tavalla todisteiden esittämisen ja kumottavissa olevien olettamien avulla”, jotta voitaisiin parantaa oikeussuojan saatavuutta tekoälyn aiheuttamien haittojen, kuten algoritmisen syrjinnän, tapauksissa (Euroopan komissio, 2022a). Peruutetun ehdotuksen 3 artiklassa todettiin erityisesti, että tuomioistuin voi määrätä suuririskisten tekoälyjärjestelmien, joiden epäillään aiheuttaneen vahinkoa, tarjoajat tai käyttöönottajat esittämään asiaankuuluvaa todistusaineistoa, jos se on tarpeen oikeudenkäyntimenettelyissä. Todisteiden esittämistä tai säilyttämistä koskevan pyynnön noudattamatta jättäminen olisi johtanut kumottavissa olevaan olettamaan, että vastaaja ei ole noudattanut velvoitteitaan. Peruutetun ehdotuksen 4 artiklan perusteluosassa todettiin, että ”[k]antajien voi olla haastavaa määrittää syy-yhteys tällaisen laiminlyönnin ja vahingon aiheuttaneen tekoälyjärjestelmän tuottaman tuotoksen tai sen puuttumisen välillä” (ibid.). Esimerkiksi syrjintä olisi sisältynyt tällaisiin vahinkoihin. Ehdotuksessa siis tekoälyasetuksen tai muiden EU:n sääntöjen mukaisten huolellisuusvelvoitteiden noudattamatta jättäminen oli ymmärrettävä tuottamukseksi, ja se olisi johtanut kumottavissa olevaan syy-yhteyttä koskevaan olettamaan tuottamustapauksessa. Jos esimerkiksi suuririskisen tekoälyjärjestelmän tarjoaja ei olisi ottanut käyttöön tekoälyasetuksen mukaista riskinhallintajärjestelmää, tuomioistuin olisi voinut olettaa, että huolellisuusvelvoitteiden laiminlyönnin ja järjestelmän aiheuttamien haittojen, kuten syrjinnän, välillä on syy-yhteys. Nämä säännöt olisivat voineet helpottaa algoritmista syrjintää koskevan *prima facie*-tapauksen osoittamista tuomioistuimessa tilanteissa, joissa vastaaja ei olisi noudattanut tekoälyasetuksen mukaisia velvoitteitaan. Kolmen edellytyksen olisi kuitenkin pitänyt täytyä: kantajan olisi pitänyt osoittaa huolellisuusvelvoitteen noudattamatta jättäminen, olisi voitava kohtuudella olettaa, että tuottamus on vaikuttanut

tekoälyjärjestelmän tuottamaan tuotokseen, ja kantajan olisi pitänyt osoittaa, että vahingon aiheutti tekoälyjärjestelmän tuottama tuotos. Yksittäisille syrjinnän uhreille näiden tekijöiden osoittaminen olisi jo itsessään aiheuttanut merkittäviä esteitä. Näin ollen syrjintäolettaa ja todistustaakan siirtämistä koskevien nykyisten sääntöjen mukauttamista on pohdittava tekoälyn ja ADM-järjestelmien konteksteissa.

Muu EU:n lainsäädäntö ja suuntaviivat

Muulla lainsäädännöllä voidaan puuttua algoritmiseen syrjintään EU:n tasolla. Esimerkiksi digipalvelusäädöksen 34 artiklassa, joka koskee erittäin suurten verkkoalustojen ja erittäin suurten verkossa toimivien hakukoneiden riskinarviointia, mainitaan syrjinnän riskit (Euroopan unioni, 2022). Ne sisältyvät myös 35 artiklaan, joka koskee riskien vähentämistä.¹² Lisäksi Euroopan komissio on äskettäin aloittanut digitaalista oikeudenmukaisuutta koskevan säädöksen julkisen kuulemisen. Tämän säädöksen tavoitteena on vahvistaa kuluttajansuojaa ja digitaalista oikeudenmukaisuutta.

EU on myös hyväksynyt tekoälyn ja ADM-järjestelmien hallinnointia koskevia tärkeitä ohjeita, joita yhdenvertaisuuselimet voivat hyödyntää esimerkiksi tehdessään poliittisia suosituksia päätöksentekijöille tai lisätessään tietoisuutta näistä järjestelmistä. Esimerkiksi vuonna 2022 vahvistettu eurooppalainen julistus digitaalisen vuosikymmenen digitaalisista oikeuksista ja periaatteista käsittelee digitaalisia oikeuksia (Euroopan komissio, 2022b). Siinä sitoudutaan ”varmistamaan, että algoritmijärjestelmät perustuvat riittäviin tietoihin, jotta vältetään syrjintä ja mahdollistetaan ihmisten suorittama valvonta aina, kun on kyse ihmisten turvallisuuteen ja perusoikeuksiin vaikuttavista tuloksista”. Myös Euroopan komission valkoisessa kirjassa (2020) käsitellään algoritmista syrjintää.

1.4.2. Syrjinnän torjunnan välineet

Vaikka nimenomaan tekoälyä koskevat säännöt ovat tärkeitä uusia välineitä yhdenvertaisuuselimille ja muille syrjintää torjuville organisaatioille, syrjinnän vastaisen lainsäädännön ja tietosuojalainsäädännön aktivointi yhdessä näiden uusien säännösten kanssa on avainasemassa algoritmisen syrjinnän torjunnassa. Tekoälyä koskevat oikeudelliset kehykset täydentävät yhdenvertaisuuslainsäädäntöä mutta eivät suinkaan korvaa sitä. Erityisesti tekoälyasetuksen vaatimukset suuririskisille tekoälyjärjestelmille tarjoavat joukon teknisiä suojatoimia, jotka tukevat – mutta eivät voi varmistaa – syrjinnän vastaisen lainsäädännön noudattamista. Tässä yhteydessä yhdenvertaisuuselimillä on tärkeä rooli nykyisten yhdenvertaisuuslainsäädännön puitteiden soveltamisessa algoritmiseen syrjintään. Ne voivat esimerkiksi antaa asiantuntevia neuvoja syrjintäolettaa koskevien sääntöjen konkreettisesta soveltamisesta ja todistustaakan siirtämisestä tapauksissa, joissa tiedot ovat epäsymmetrisiä. Ne voivat myös olla tärkeässä asemassa, kun algoritmeista aiheutuvia haittoja

12. Katso yhdenvertaisuuselinten mahdollisuuksia käsitellä tekoälyyn liittyvää verkkosyrjintää, kuten vihapuhetta, esimerkiksi seuraavasta: Equinet (2018), ”Extending the agenda. Equality bodies addressing hate speech”, saatavilla osoitteessa <https://equineteurope.org/extending-the-agenda-equality-bodies-addressing-hate-speech/>, viitattu 7. marraskuuta 2025; Facing Facts Network / CEJI (2022), ”Current activities and gaps in hate speech responses: a mapping report for the Facing Facts Network”, saatavilla osoitteessa www.facingfacts.eu/hate-speech-report/, viitattu 7. marraskuuta 2025.

arvioidaan oikeudellisesti välittömän, välillisen ja intersektionaalisen syrjinnän kannalta, tai neuvoa esimerkiksi kysymyksissä liittyen siihen, miten poikkeuksia ja objektiivisia perusteluja voidaan soveltaa tässä kontekstissa.

Euroopan neuvosto

Euroopan neuvoston tasolla useat välineet kieltävät syrjinnän, ja niitä voidaan soveltaa algoritmiseen syrjintään.

Tärkein väline on Euroopan ihmisoikeussopimuksen 14 artikla, joka kieltää syrjinnän. Siinä todetaan, että "[t]ässä yleissopimuksessa tunnustetuista oikeuksista ja vapauksista nauttiminen taataan ilman minkäänlaista sukupuoleen, rotuun, ihonväriin, kieleen, uskontoon, poliittisiin tai muihin mielipiteisiin, kansalliseen tai yhteiskunnalliseen alkuperään, kansalliseen vähemmistöön kuulumiseen, varallisuuteen, syntyperään tai muuhun asemaan perustuvaa syrjintää". Tämä ei ole tyhjentävä luettelo kielletyistä syrjinnän muodoista, mutta sitä voidaan käyttää tekoäly- ja ADM-järjestelmien syrjivien vaikutusten torjuntaan, jos jotakin yleissopimuksessa lueteltua perusoikeutta on loukattu.

Tätä täydentää yleissopimuksen kahdennentoista pöytäkirjan 1 artiklan 1 kohta, jossa määrätään kaikkinaisen syrjinnän kiellosta: "Laissa tunnustetuista oikeuksista nauttiminen taataan ilman minkäänlaista sukupuoleen, rotuun, ihonväriin, kieleen, uskontoon, poliittisiin tai muihin mielipiteisiin, kansalliseen tai yhteiskunnalliseen alkuperään, kansalliseen vähemmistöön kuulumiseen, varallisuuteen, syntyperään tai muuhun asemaan perustuvaa syrjintää." Lisäksi 1 artiklan 2 kohdassa todetaan, että "[k]etään ei saa syrjiä viranomaistoiminnassa 1 kappaleessa tarkoitetuilla perusteilla". Tätä syrjintäkieltoa sovelletaan – riippumatta siitä, rikotaanko jotain muuta perusoikeutta – julkishallintoihin, jotka käyttävät syrjiviä tekoäly- ja ADM-järjestelmiä kahdennentoista pöytäkirjan ratifioineissa maissa. Tätä syrjinnän vastaista kehystä täydentää työelämän osalta Euroopan sosiaalisen peruskirjan 20 artikla, joka koskee oikeutta tasa-arvoisiin mahdollisuuksiin ja tasa-arvoiseen kohteluun työhön ja ammattiin liittyvissä kysymyksissä ilman sukupuoleen perustuvaa syrjintää.

Muita asiaankuuluvia välineitä ovat muun muassa Euroopan neuvoston yleissopimus naisiin kohdistuvan väkivallan ja perheväkivallan ehkäisemisestä ja torjumisesta (Istanbulin yleissopimus), jonka 1 artiklan 1 kohdan b alakohdalla pyritään "edistämään kaikkien naisten syrjinnän muotojen poistamista sekä naisten ja miesten välistä tosiasiallista tasa-arvoa, myös voimaannuttamalla naisia". Yleissopimuksessa todetaan myös, että naisiin kohdistuva väkivalta "tarkoittaa naisiin kohdistuvaa ihmisoikeusloukkausta ja yhtä syrjinnän muotoa". Kansallisten vähemmistöjen suojelua koskevan puiteyleissopimuksen 4 artiklassa puolestaan todetaan, että "kaikki syrjintä kansalliseen vähemmistöön kuulumiseen perusteella on kielletty". Tämä sisältää "[henkilön] joutu[misen] syrjinnän, vihamielisyyden tai väkivallan tai niiden uhan kohteeksi etnisen, sivistyksellisen, kielellisen tai uskonnollisen identiteettinsä johdosta" (6 artikla). Alueellisia kieliä tai vähemmistökieliä koskevan eurooppalaisen peruskirjan 7 artiklan 2 kohdassa suojellaan vastaavasti henkilöitä syrjinnältä, joka perustuu "alueellisiin kieliin tai vähemmistökieliin".

Näiden perusvälineiden lisäksi algoritmisen syrjinnän torjunnassa voidaan hyödyntää myös muita tärkeitä asiakirjoja. Esimerkiksi Euroopan neuvoston sukupuolten

tasa-arvostrategiassa 2024–2029 tunnustetaan ”tekoälyn erityinen vaikutus sukupuolten tasa-arvoon ja naisten oikeuksiin” ja kehoitetaan ”jäsenvaltioita puuttumaan sukupuoleen perustuvaan ja intersektionaaliseen algoritmiseen syrjintään ihmisoikeuksiin perustuvilla ja monivaikutteisilla sukupuolten yhdenvertaisuutta ja syrjimättömyyttä koskevilla strategioilla [ja] panemalla täytäntöön tekoälyä ja sukupuolten yhdenvertaisuutta koskevat uudet standardit” (Euroopan neuvosto, 2024c). Tämä luo jäsenvaltioille toimeksiannon tarkistaa lainsäädäntöään, jotta algoritmiseen syrjintään liittyviä puutteita voidaan korjata.

Ministerikomitean jäsenvaltioille antamassa suosituksessa CM/Rec(2019)1 seksismin ehkäisemisestä ja torjunnasta tunnustetaan, että ”internet on luonut uuden ulottuvuuden seksismin ja erityisesti seksistisen vihapuheen ilmaisemiseen ja välttämiseen suurelle yleisölle, vaikka seksismin juuret eivät olekaan teknologiassa vaan sitkeässä sukupuolten välisessä eriarvoisuudessa” (Euroopan neuvosto, 2019b). Suosituksessa todetaan, että ”tekoäly asettaa erityisiä haasteita sukupuolten väliselle yhdenvertaisuudelle ja sukupuolistereotypioille” ja että ”[a]lgoritmien käyttö voi välittää ja vahvistaa olemassa olevia sukupuolistereotypioita ja siten edistää seksismin jatkumista”. Siinä tunnustetaan tekoäly- ja ADM-järjestelmien rooli ”naisten kehojen, puheen ja aktivismin tarkastelun syventämisessä” erityisesti verkossa. Suosituksessa jäsenvaltioita kehoitetaan ”[s]isällyttämään sukupuolten välistä yhdenvertaisuutta koskeva näkökulma kaikkiin tekoälyyn liittyviin poliittisiin päätöksiin, ohjelmiin ja tutkimuksiin, jotta voidaan välttää teknologian mahdollisesti aiheuttamat seksismi- ja sukupuolistereotypiariskit sekä selvittää, miten tekoäly voisi auttaa sukupuolten välisen eriarvoisuuden vähentämisessä ja seksismin poistamisessa”. Siinä suositellaan ”naisten ja tyttöjen osallistumisen lisäämistä tieto- ja teknologia-alalla” ja vaaditaan, että ”tietoon perustuvien välineiden ja algoritmien suunnittelussa... otetaan huomioon sukupuoleen perustuva dynamiikka”. Lisäksi ”[n]äihin ongelmiin liittyvää avoimuutta on parannettava, tietoisuutta massadatan mahdollisesta sukupuoleen perustuvasta vinoumasta on lisättävä ja ratkaisuja vastuuvollisuuden kehittämiseksi on esitettävä”. Kaiken kaikkiaan tämä suositus on tärkeä oikeudellinen väline algoritmisen syrjinnän torjunnassa.

Ministerikomitean jäsenvaltioille antama suositus CM/Rec(2022)17 maahanmuuttaja-, pakolais- ja turvapaikanhakijanaisten ja -tyttöjen oikeuksien suojelusta sisältää tekoälyä, ADM-järjestelmiä ja tietosuojaa koskevan osion. Suosituksen mukaan ”julkisen tai yksityisen sektorin tai palveluntarjoajien ja alihankkijoiden tekoälyn ja automaattisen päätöksenteon järjestelmien suunnittelun, kehittämisen ja soveltamisen tulee olla syrjimätöntä, yksityisyyden suojaa koskevien periaatteiden mukaista ja avointa, ja sillä tulee olla selkeät hallintomekanismit” rajavalvonnan ja muuttoliikkeen hallinnan konteksteissa (Euroopan neuvosto, 2022c). Se kannustaa valtioita varmistamaan, että ihmisoikeusvaikutusten arviointi tehdään ennen tekoälyn ja ADM-järjestelmien käyttöönottoa muuttoliikkeen alalla. Suosituksessa edellytetään myös, että kansalaisjärjestöt otetaan mukaan keskusteluihin, jotka koskevat maahanmuuttaja-, pakolais- ja turvapaikanhakijanaisten ja -tyttöjen asemaan vaikuttavan uuden teknologian kehittämistä ja käyttöönottoa (ibid.).

Euroopan neuvoston ministerikomitean suuntaviivat yhdenvertaisuuden ylläpitämisestä sekä syrjinnän ja vihan torjumisesta covid-19-pandemian ja vastaavien kriisien aikana tulevaisuudessa sisältävät useita kohtia, joissa käsitellään digitalisaation,

tekoälyn ja kontaktien jäljitykseen käytettävien teknologioiden "haavoittuviin ryhmiin" kohdistuvaa syrjivää potentiaalia, vaikka painopiste onkin digitaalisessa syrjäytymisessä ja käyttömahdollisuuksien puutteessa (Euroopan neuvosto, 2021c). Suosituksessa CM/Rec(2022)16 vihapuheen torjumisesta ja suosituksessa CM/Rec(2024)4 viharikosten torjumisesta jäsenvaltioita kehoitetaan "suojelemaan ihmisoikeuksia ja perusvapauksia digitaalisessa ympäristössä". Niissä korostetaan internetissä toimivien välittäjien roolia ja vastuuta vihapuheen levittämisessä, tuodaan esille digitaalisten todisteiden ja datan merkitystä viharikosten tutkinnassa ja tunnustetaan "ensisijaisesti digitaalisten kanavien ja verkkoviestinnän välityksellä toimivien ääriliikkeiden" rooli viharikoksissa (Euroopan neuvosto, 2022d). Nämä suositukset eivät kuitenkaan puutu täydellisesti siihen, miten sosiaalisen median algoritmiset suositusjärjestelmät edistävät äärimmäistä polarisoitumista verkossa ja siten syrjivien stereotyyppien ja ennakkoluulojen syntymistä käyttäjien keskuudessa.

Euroopan neuvoston yleissopimuksessa lasten suojelemisesta seksuaalista riistoa ja seksuaalista hyväksikäyttöä vastaan (Lanzaroten yleissopimus) tunnustetaan "sekä lasten että rikoksentehtäjäiden lisäänty[nyt] informaatio- ja viestintäteknologian käyt[tö]" ja todetaan, että digitaalinen teknologia helpottaa lasten hyväksikäyttöä (Euroopan neuvosto, 2007). Vuonna 2010 hyväksyttyä ministerikomitean suositusta CM/Rec(2010)5 toimista seksuaaliseen suuntautumiseen tai sukupuoli-identiteettiin perustuvan syrjinnän torjumiseksi voidaan käyttää monialaisesti algoritmiseen syrjintään puuttumiseen, mutta se ei sisällä säännöksiä, joissa käsiteltäisiin nimenomaisesti tekoälyn ja ADM-järjestelmien syrjiviä vaikutuksia, vaikka yhä useammat todisteet osoittavat, että nämä teknologiat lisäävät ennakkoluuloja ja syrjäytymistä sekä verkossa että sen ulkopuolella (Euroopan neuvosto, 2010). Sama pätee olemassa oleviin suosituksiin, joilla kielletään romanien ja kiertolaisten syrjintä, esimerkiksi ministerikomitean suositukseen CM/Rec(2024)1 romani- ja kiertolaisnaisten ja -tyttöjen yhdenvertaisuudesta (Euroopan neuvosto, 2024d).

Euroopan rasismien ja suvaitsemattomuuden vastaisen komission yleiset poliittiset suositukset tarjoavat myös hyödyllisiä oikeudellisia resursseja algoritmiseen syrjintään puuttumiseen, erityisesti tekoälyn ja ADM-järjestelmien alakohtaisen soveltamisen näkökulmasta:

- ▶ Rasistisen, muukalaisvihamielisen ja antisemitistisen materiaalin levittämisen torjumista internetissä koskeva suositus nro 6 voi auttaa torjumaan rasistista sisältöä, vihapuhetta, verkkohäirintää ja ääriliikkeitä, joita algoritmiset suositusjärjestelmät ruokkivat (Euroopan rasismien ja suvaitsemattomuuden vastainen komissio, 2000).
- ▶ Suositus nro 7 kansallisesta lainsäädännöstä rasismien ja rotusyrjinnän torjumiseksi tarjoaa mielenkiintoisia menettelyllisiä keinoja algoritmisen syrjinnän torjumiseksi rakenteellisella tasolla:

Laissa olisi säädettävä, että organisaatioilla, kuten yhdistyksillä, ammattiyhdistyksillä ja muilla oikeushenkilöillä, joilla on kansallisessa lainsäädännössä vahvistettujen kriteerien mukaisesti oikeutettu intressi torjua rasismia ja rotusyrjintää, on oikeus nostaa siviilioikeudellisia kanteita, puuttua hallinnollisiin asioihin tai tehdä rikosilmoituksia, vaikka tarkkaa uhria ei mainittaisikaan (Euroopan rasismien ja suvaitsemattomuuden vastainen komissio, 2017: 8)

...

[Tämä on] olennaisen tärkeää käsiteltäessä syrjintätapauksia, joissa uhria on vaikea tunnistaa, tai tapauksia, jotka koskevat määrittelemätöntä määrää uhreja (ibid.: 22).

► Rasismin torjuntaa terrorismin vastaisessa taistelussa koskevassa suosituksessa nro 8 jäsenvaltioita kehoitetaan

varmistamaan, että erityisesti seuraavia aloja koskeva lainsäädäntö ja asetukset – tai niiden täytäntöönpano – eivät aiheuta syrjintää:

- maiden lainvalvontaviranomaisten ja rajavaltion henkilöstön suorittamat tarkastukset
- hallinnollinen säilöönnotto ja tutkintavankeus
- ...
- oikeudenmukainen oikeudenkäynti ja rikosprosessi
- henkilötietojen suoja
- yksityis- ja perhe-elämän suoja
- maasta poistaminen, rikoksen johdosta tapahtuva luovuttaminen, maasta karkottaminen ja palautuskiellon periaate
- viisumien myöntäminen
- oleskelu- ja työluvut sekä perheen yhdistäminen
- kansalaisuuden myöntäminen ja kumoaminen

(Euroopan rasismin ja suvaitsemattomuuden vastainen komissio, 2004).

Koska tekoälyä ja ADM-järjestelmiä käytetään yhä enemmän näillä aloilla ja erityisesti lainvalvonnan helpottamiseksi, suositus tarjoaa hyödyllisen kehyksen niiden syrjivään potentiaaliin puuttumiseksi, vaikka se ei itsessään suoranaisesti koske näitä teknologioita.

- Suositusta nro 10 rasismin ja rotusyrjinnän torjumisesta kouluopetuksessa ja sen avulla sovelletaan myös silloin, kun tekoäly- ja ADM-järjestelmiä käytetään esimerkiksi valintamenettelyissä (Euroopan rasismin ja suvaitsemattomuuden vastainen komissio, 2006).
- Myös suositus nro 11 rasismin ja rotusyrjinnän torjunnasta poliisitoiminnassa eli sektorilla, jolla tekoälyä ja ADM-järjestelmiä käytetään esimerkiksi kasvontunnistus- ja yhteensovistusteknologiassa sekä rikosten ennustamiseen ja valvontaan tarkoitetuissa järjestelmissä, on merkityksellinen (Euroopan rasismin ja suvaitsemattomuuden vastainen komissio, 2007).
- Suositus nro 14 rasismin ja rotusyrjinnän torjumisesta työelämässä tarjoaa hyödyllisiä suuntaviivoja alalla, jolla käytetään yhä enemmän ansioluetteloiden seulonnan ja rekrytoinnin tukivälineitä (Euroopan rasismin ja suvaitsemattomuuden vastainen komissio, 2012).
- Suosituksessa nro 15 vihapuheen torjunnasta vaaditaan, että jäsenvaltioiden on "varmistettava, että [vihapuheeseen liittyvien] rikosten soveltamisala määritellään tavalla, joka mahdollistaa niiden soveltamisen teknologian kehityksen mukana" (Euroopan rasismin ja suvaitsemattomuuden vastainen komissio, 2015).

Tätä kehystä täydennetään muilla suosituksilla, joissa keskitytään tiettyihin syrjinnän muotoihin, kuten rasismiin, muukalaisvihaan, romanivastaisuuteen, antisemitismiin, muslimivihaan sekä luvattomasti maassa oleskeleviin maahanmuuttajiin ja lesboihin, homoihin, biseksuaaleihin ja transsukupuolisiin ja intersukupuolisiin henkilöihin (LHBTI) kohdistuvaan syrjintään.¹³

Euroopan unioni

Euroopan unionin tasolla syrjimättömyyden ja sukupuolten välisen tasa-arvon periaatteet taataan SEU:n unionin arvoja koskevassa 2 artiklassa ja sisämarkkinoita koskevassa 3 artiklassa sekä Euroopan unionin toiminnasta tehdyn sopimuksen (SEUT) yhdenvertaisuuden ja syrjimättömyyden valtavirtaistamista koskevassa 8 ja 10 artiklassa, SEUT:n 19 artiklassa, jossa EU:lle annetaan toimeksianto hyväksyä syrjinnän vastaista lainsäädäntöä, ja SEUT:n palkkatasa-arvoa koskevassa 157 artiklassa. EU:n perusoikeuskirjan 21 artiklan 1 kohdan mukaan "[k]ielletään kaikenlainen syrjintä, joka perustuu sukupuoleen, rotuun, ihonväriin tai etniseen taikka yhteiskunnalliseen alkuperään, geneettisiin ominaisuuksiin, kieleen, uskontoon tai vakaumukseen, poliittisiin tai muihin mielipiteisiin, kansalliseen vähemmistöön kuulumiseen, varallisuuteen, syntyperään, vammaisuuteen, ikään tai sukupuoliseen suuntautumiseen tai muuhun sellaiseen seikkaan". Perusoikeuskirjan 23 artiklassa todetaan, että "[n]aisten ja miesten välinen tasa-arvo on varmistettava kaikilla aloilla työelämä ja palkkaus mukaan lukien".

Useilla eri direktiiveillä vahvistetaan kattava syrjinnän vastaisten sääntöjen verkosto. Tällaisia ovat esimerkiksi seuraavat:

- neuvoston direktiivi 2000/43/EY rodusta tai etnisestä alkuperästä riippumattoman yhdenvertaisen kohtelun periaatteen täytäntöönpanosta
- neuvoston direktiivi 2000/78/EY yhdenvertaista kohtelua työssä ja ammatissa koskevista yleisistä puitteista
- neuvoston direktiivi 2004/113/EY miesten ja naisten yhdenvertaisen kohtelun periaatteen täytäntöönpanosta tavaroiden ja palvelujen saatavuuden ja tarjonnan alalla
- Euroopan parlamentin ja neuvoston direktiivi 2006/54/EY miesten ja naisten yhtäläisten mahdollisuuksien ja yhdenvertaisen kohtelun periaatteen täytäntöönpanosta työhön ja ammattiin liittyvissä asioissa (uudelleenlaadittu toisinto)
- Euroopan parlamentin ja neuvoston direktiivi 2010/41/EU miesten ja naisten tasa-arvoisen kohtelun periaatteen soveltamisesta itsenäisiin ammatinharjoittajiin
- neuvoston direktiivi 79/7/ETY miesten ja naisten tasa-arvoisen kohtelun periaatteen asteittaisesta toteuttamisesta sosiaaliturvaa koskevissa kysymyksissä.

Näitä voidaan soveltaa algoritmiseen syrjintään, kun syrjiviä tekoäly- ja ADM-järjestelmiä käytetään niiden olennaisella soveltamisalalla. Soveltamisalassa on kuitenkin aukkoja, jotka liittyvät esimerkiksi ikään, seksuaaliseen suuntautumiseen,

13. Katso kaikki suositukset osoitteesta www.coe.int/en/web/european-commission-against-racism-and-intolerance/ecri-standards, viitattu 9. marraskuuta 2025.

vammaisuuteen ja uskontoon tai vakaumukseen perustuvaan syrjintään tavaroiden ja palvelujen ostamisen yhteydessä.

Yhdenvertaisuuselimä koskevat Euroopan unionin direktiivit

Kahdella direktiivillä on hiljattain uudistettu sellaisten yhdenvertaisuuselinten toimeksiantoja, jotka ovat keskeisiä toimijoita algoritmisen syrjinnän torjunnassa:

- ▶ direktiivi 2024/1499 normeista tasa-arvoelimille, jotka toimivat asioissa, jotka koskevat henkilöiden rodusta tai etnisestä alkuperästä riippumatonta yhdenvertaista kohtelua, uskonnosta tai vakaumuksesta, vammaisuudesta, iästä tai seksuaalisesta suuntautumisesta riippumatonta yhdenvertaista kohtelua työssä ja ammatissa ja naisten ja miesten yhdenvertaista kohtelua sosiaaliturvaa koskevista asioista ja tavaroiden ja palvelujen saatavuuden ja tarjonnan osalta
- ▶ direktiivi 2024/1500 tasa-arvoelimä koskevista normeista naisten ja miesten yhdenvertaiseen kohteluun ja yhtäläisiin mahdollisuuksiin työssä ja ammatissa liittyvissä asioissa.

EU:n jäsenvaltioiden on kesäkuuhun 2026 mennessä mukautettava kansallinen lainsäädäntönsä kahden yhdenvertaisuuselimä koskevan EU:n direktiivin säännösten mukaiseksi. Automaattisten järjestelmien ja tekoälyn osalta näissä direktiiveissä säädetään, että "tasa-arvoelimillä olisi erityisesti oltava käytettävissään sopivat henkilöstö- ja tekniset resurssit. Kyseisten resurssien olisi etenkin mahdollistettava se, että tasa-arvoelimet käyttävät automatisoituja järjestelmiä työssään ja toisaalta arvioivat sitä, ovatko kyseiset järjestelmät syrjimättömyysääntöjen mukaisia".

1.4.3. Tietosuojavälineet

Tietosuojalainsäädäntö on toinen osa lainsäädännöllistä palapeliä, jota yhdenvertaisuuselinten ja muiden eriarvoisuutta Euroopassa torjuvien organisaatioiden on käytettävä algoritmisen syrjinnän torjumiseksi.

Euroopan neuvosto

Euroopan neuvoston tasolla tietosuojaa koskevaa lainsäädäntökehystä voidaan hyödyntää algoritmiseen syrjintään puuttumisessa. Uudistetussa yleissopimuksessa yksilöiden suojelusta henkilötietojen automaattisessa tietojenkäsittelyssä (yleissopimus 108+) ja erityisesti sen 6 artiklassa, jossa rajoitetaan erityisten tietoryhmien käyttöä, kehoitetaan jäsenvaltioita ottamaan käyttöön "suojatoimia... suojautumiseen riskeiltä, joita arkaluonteisten tietojen käsittely voi aiheuttaa rekisteröidyn eduille, oikeuksille ja perusvapauksille, erityisesti syrjinnän riskiltä" (Euroopan neuvosto, 2018). Tekoälyä koskevan puiteyleissopimuksen yksityisyyttä ja henkilötietojen suojaa koskevassa 11 artiklassa nämä rajoitukset laajennetaan "tekoälyjärjestelmien elinkaareen kuuluiin toimiin". Jäsenvaltioita pyydetään erityisesti varmistamaan, että "henkilöiden oikeutta yksityisyyteen ja heidän henkilötietojaan suojellaan, myös sovellettavien kansallisten ja kansainvälisten lakien, normien ja kehysten avulla" ja että "henkilöitä varten on otettu käyttöön tehokkaat takeet ja suojatoimet sovellettavien kansallisten ja kansainvälisten oikeudellisten velvoitteiden mukaisesti". Yleissopimusta 108+

sovelletaan sen tultua voimaan myös kansallisen turvallisuuden ja puolustuksen aloilla, joten sen soveltamisala on laajempi kuin yleisen tietosuoja-asetuksen.

Tätä kehystä täydentävät tekoälyä ja tietosuoja koskevat suuntaviivat (Guidelines on artificial intelligence and data protection), jotka yksilöiden suojelusta henkilö-tietojen automaattisessa tietojenkäsittelyssä tehdyn yleissopimuksen (yleissopimus 108) neuvoo-antava komitea antoi vuonna 2019 (Euroopan neuvosto, 2019c). Siinä korostetaan, että "[k]aikissa käsittelyvaiheissa, myös tiedonkeruussa, tekoälyn kehittäjien, valmistajien ja palveluntarjoajien on käytettävä sisäänrakennettujen ihmisoikeuksien lähestymistapaa ja vältettävä mahdollisia vinoumia, myös tahattomia tai piilotettuja, sekä syrjinnän riskiä tai muita haittavaikutuksia rekisteröityjen ihmisoikeuksiin ja perusvapauksiin". Siinä kannustetaan myös "[y]hteistyöhön ... tietosuoja- ja valvontaviranomaisten ja muiden sellaisten elinten välillä, joilla on tekoälyyn liittyvää toimivaltaa, kuten kuluttajansuojasta, kilpailusta, syrjinnän torjunnasta, eri sektorien sääntelystä ja median sääntelystä vastaavat valvontaviranomaiset".

Lisäksi vuonna 2021 hyväksytyissä kasvontunnistusta koskevissa suuntaviivoissa säädetään, että "[k]asvontunnistuksen käyttö yksinomaan henkilön ihonvärin, uskonnollisen tai muun vakaumuksen, sukupuolen, rodun tai etnisen alkuperän, iän, terveydentilan tai yhteiskunnallisen tilanteen määrittämiseksi on kiellettävä, ellei laissa säädetä asianmukaisista suojoitoista syrjintäriskin välttämiseksi" (Euroopan neuvosto, 2021d). Suuntaviivat tarjoavat myös ohjeita datajoukkojen edustavuudesta kehittäjille, valmistajille ja palveluntarjoajille. Suuntaviivojen suositusten perustana on yleissopimuksessa 108+ oleva 5 artikla tietojen tarkkuudesta, ja niissä todetaan, että kehittäjien, valmistajien ja käyttäjien "on vältettävä virheellisiä merkintöjä, testattava järjestelmiään riittävästi ja tunnistettava ja poistettava tarkkuuseroja, jotka liittyvät erityisesti ihonvärin, iän ja sukupuolen demografisiin vaihteluihin, ja vältettävä näin tahatonta syrjintää". Suuntaviivoissa pyydetään myös, että lainvalvontaviranomaiset "harkitessaan kasvontunnistusteknologian käyttöönottoa valvomattomissa ympäristöissä... ottavat huomioon eri perusoikeuksiin, kuten tietosuojaan, yksityisyyteen, sananvapauteen, kokoontumisvapauteen ja liikkumisvapauteen tai syrjintäkieltoon, kohdistuvan riskin riippuen mahdollisista käyttötavoista eri paikoissa" erityisesti tietosuojan vaikutustenarviointien avulla.

Euroopan unioni

Euroopan unionin tasolla useilla yleisen tietosuoja-asetuksen määräyksillä voi olla tärkeä rooli tekoäly- ja ADM-järjestelmiin liittyvän syrjinnän käsittelyssä (Euroopan unioni, 2016a). Ensinnäkin arkaluonteisia henkilötietoryhmiä koskeva 9 artikla estää periaatteessa tarjoajia ja käyttöönottajia käyttämästä tällaista dataa muuttujina (esimerkiksi arvot, syötemuuttujat tai riskitekijät) tekoäly- ja ADM-järjestelmissä. Kuten aiemmin todettiin, tekoälyasetuksen 10 artiklan 5 kohta sallii kuitenkin tällaisen käytön yhdessä yleisen tietosuoja-asetuksen 9 artiklan 1 kohdan g alakohdan kanssa, kun tavoitteena on algoritmisen syrjinnän tunnistaminen ja vähentäminen. Tekoälyasetuksen johdanto-osan kappaleessa 70 selvennetäänkin seuraavaa:

Jotta voitaisiin suojella muita tekoälyjärjestelmissä esiintyvistä vinoutumisesta mahdollisesti johtuvalta syrjinnältä, tarjoajien olisi suuririskisten tekoälyjärjestelmien vinoutumien havaitsemisen ja korjaamisen varmistamiseksi voitava, poikkeuksellisesti

ja siinä määrin kuin se on ehdottoman välttämätöntä sekä soveltaen luonnollisten henkilöiden perusoikeuksia ja -vapauksia koskevia asianmukaisia suoja-toimia,... käsitellä myös erityisiä henkilötietoryhmiä [yleisen tietosuojaa-asetuksen 9 artiklan 2 kohdan g alakohdassa] tarkoitetun tärkeän yleisen edun nimissä.

Lisäksi yleisen tietosuojaa-asetuksen 22 artiklassa annetaan oikeus olla joutumatta täysin automaattisten päätösten kohteeksi. Erityisesti todetaan, että "[r]ekisteröidyillä on oikeus olla joutumatta sellaisen päätöksen kohteeksi, joka perustuu pelkästään automaattiseen käsittelyyn, kuten profilointiin, ja jolla on häntä koskevia oikeusvaikutuksia tai joka vaikuttaa häneen vastaavalla tavalla merkittävästi". Tätä säännöstä on käytetty syrjivän luottopisteytyksen haastamiseen asioissa SCHUFA Holding ja Dun & Bradstreet Austria (katso kohta 1.4.4.). Vaikka yleisen tietosuojaa-asetuksen johdanto-osan kappaleessa 71 myönnetään rekisteröidyille oikeus "saada selvitys... arvioinnin jälkeen tehdystä päätöksestä", kun päätös on tehty täysin ADM-järjestelmän avulla, pitkään vallitsi oikeudellinen epävarmuus siitä, tarjoaako yleinen tietosuojaa-asetus oikeuden selvitykseen, millainen tällainen oikeus mahdollisesti käytännössä on ja millaiset rajat sillä on (Wachter, Mittelstadt ja Floridi, 2017; Malgieri ja Comandé, 2017; Selbst ja Powles, 2017). Yleisen tietosuojaa-asetuksen mukaan rekisteröidyillä on oikeus saada "tiedot, jotka ovat tarpeen asianmukaisen ja läpinäkyvän käsittelyn takaamiseksi" (13 artiklan 2 kohta ja 14 artiklan 2 kohta), mukaan luettuna "automaattisen päätöksenteon, muun muassa 22 artiklan 1 ja 4 kohdassa tarkoitetun profiloinnin olemassaolo, sekä ainakin näissä tapauksissa merkitykselliset tiedot käsittelyyn liittyvästä logiikasta samoin kuin kyseisen käsittelyn merkittävyys ja mahdolliset seuraukset rekisteröidyille" (13 artiklan 2 kohdan f alakohta, 14 artiklan 2 kohdan g alakohta ja 15 artiklan 1 kohdan h alakohta). Dun & Bradstreet Austria -asian päätöksessä (katso kohta 1.4.4.) Euroopan unionin tuomioistuin vahvisti yleisen tietosuojaa-asetuksen 15 artiklan 1 kohdan h alakohtaan sisältyvän oikeuden saada selvitys olemassaolon ja laajuuden. Tämä tarkoittaa oikeutta saada merkityksellisiä tietoja siitä, miten päätös on tehty, ja tarjoaa perustan haastaa automaattiset päätökset yleisen tietosuojaa-asetuksen 22 artiklan 3 kohdan perusteella. Yleisen tietosuojaa-asetuksen 22 artiklan 1 kohdassa todetaan, että "[r]ekisteröidyillä on oikeus olla joutumatta sellaisen päätöksen kohteeksi, joka perustuu pelkästään automaattiseen käsittelyyn, kuten profilointiin, ja jolla on häntä koskevia oikeusvaikutuksia tai joka vaikuttaa häneen vastaavalla tavalla merkittävästi". Tätä oikeutta ei kuitenkaan sovelleta, jos automaattinen käsittely tai profilointi on "välttämätön rekisteröidyn ja rekisterinpitäjän välisen sopimuksen tekemistä tai täytäntöönpanoa varten" tai jos rekisteröity on antanut siihen "nimenomaisen suostumuksen" yleisen tietosuojaa-asetuksen 22 artiklan 2 kohdan a ja c alakohdan mukaisesti. Tällaisessa tapauksessa yleisen tietosuojaa-asetuksen 22 artiklan 3 kohdassa edellytetään, että "rekisterinpitäjän on toteutettava asianmukaiset toimenpiteet rekisteröidyn oikeuksien ja vapauksien sekä oikeutettujen etujen suojaamiseksi; tämä koskee vähintään oikeutta vaatia, että tiedot käsittelee rekisterinpitäjän puolesta luonnollinen henkilö, sekä oikeutta esittää kantansa ja riitauttaa päätös". Unionin tuomioistuimen hiljattain määrittelemä oikeus saada selvitys voi helpottaa oikeussuojan saatavuutta ainakin tapauksissa, joissa on kyse profilointiin perustuvasta algoritmista syrjinnästä tai täysin automaattisista ADM-prosesseista.

Rikosoikeuden alalla sovelletaan direktiiviä 2016/680 luonnollisten henkilöiden suojelusta toimivaltaisten viranomaisten suorittamassa henkilötietojen käsittelyssä

rikosten ennalta estämistä, tutkimista, paljastamista tai rikoksiin liittyviä syytetoimia tai rikosoikeudellisten seuraamusten täytäntöönpanoa varten sekä näiden tietojen vapaasta liikkuvuudesta (Euroopan unioni, 2016b). Se tarjoaa rajoitetumpia takuita, jotka eivät välttämättä riitä haastamaan algoritmista syrjintää kyseisillä käyttöalueilla.

1.4.4. Euroopan ihmisoikeustuomioistuimen ja Euroopan unionin tuomioistuimen oikeuskäytäntö

Euroopan ihmisoikeustuomioistuin

Asian *Glukhin v. Venäjä* päätöksessä vuonna 2023 Euroopan ihmisoikeustuomioistuin arvioi poliisin käyttämien kasvontunnistusteknologioiden laillisuutta (Euroopan ihmisoikeustuomioistuin, 2023). Tässä tapauksessa henkilö pidätettiin sen jälkeen, kun hän oli julkaissut sosiaalisessa mediassa valokuvia ja videon Moskovan metrossa yksin järjestämästään rauhanomaisesta mielenosoituksesta. Poliisi havaitsi kuvat ja videon rutiininomaisen internetseurannan aikana. Poliisi käytti henkilön tunnistukseen julkisen tilan valvontakameroiden kuvia yhdistettynä biometrisiin tietoihin, mikä tässä tapauksessa tarkoitti kasvontunnistusteknologiaa. Tuomioistuin totesi, että kantajan sananvapautta Euroopan ihmisoikeussopimuksen 10 artiklan mukaisesti ja oikeutta nauttia yksityis- ja perhe-elämän kunnioitusta Euroopan ihmisoikeussopimuksen 8 artiklan mukaisesti oli loukattu. Tuomioistuin totesi erityisesti, että ”erittäin tunkeilevan kasvontunnistusteknologian käyttö, kun kantaja käyttää ihmisoikeussopimuksessa turvattua oikeuttaan sananvapauteen, on ristiriidassa niiden oikeusvaltioperiaatteen mukaisten demokraattisen yhteiskunnan ihanteiden ja arvojen kanssa, joita sopimus on tarkoitettu ylläpitämään ja edistämään”. Ottaen huomioon sen, että ”kantajan oli vaikea todistaa väitteitään, koska kansallisessa lainsäädännössä ei säädetty kasvontunnistusteknologian käytöstä tehtävästä virallisesta kirjauksesta tai ilmoituksesta”, tuomioistuin tunnusti myös, että tällaisten teknologioiden käytön avoimuuden puute vaikeuttaa todisteiden esittämistä ja tällaisten esteiden lieventämistä (ibid., kohta 72).

Euroopan unionin tuomioistuin

Belgian esittämä ennakkoratkaisupyyntö EU:n matkustajarekisteritietoja koskevan direktiivin laajasta täytäntöönpanosta, mukaan lukien siitä, miten kansallisten viranomaisten olisi toteutettava tietojenkäsittelyjärjestelmän automatisointi, johti päätökseen asiassa C-817/19 *Ligue des droits humains ASBL v. Conseil des ministres* (Euroopan unionin tuomioistuin, 2022). Tapauksessa keskityttiin koneoppimisalgoritmien avulla tapahtuvan automaattisen tietojenkäsittelyn ja syrjimättömyyden väliseen jännitteeseen. Tuomioistuin korosti automaattista tietojenkäsittelyä koskevia suojatoimia, jotka estävät itseoppivien tekoälyalgoritmien käytön syrjivien tulosten estämiseksi. Terrorismiuhkien arvioinnissa käytettävät ennalta määritellyt kriteerit eivät myöskään voi perustua ”henkilön rotuun tai etniseen alkuperään, uskontoon tai vakaumukseen, poliittisiin tai muihin mielipiteisiin, ammattiliiton jäsenyyteen eivätkä terveydentilaa, sukupuolielämää tai sukupuolista suuntautuneisuutta koskeviin tietoihin” (ibid.).

Tuomioistuin luokitteli asian C-634/21 *SCHUFA Holding* tuomiossaan luottoihin liittyvän riskipisteytyksen muodossa tapahtuvan profiloinnin yleisen tietosuojasetuksen 22 artiklan soveltamisalaan kuuluvaksi automatisoiduksi yksittäispäätökseksi (Euroopan unionin tuomioistuin, 2023). Tähän tuomioon johtaneessa asiassa rahoituslaitos kieltäytyi myöntämästä kantajalle lainaa saksalaisen SCHUFA Holding -luottopisteytysyrityksen antaman riskipisteytyksen perusteella. Kysymys oli siitä, oliko ”kyseessä olevan kaltaisen todennäköisyysarvon määrittäminen yleisen tietosuojasetuksen 22 artiklan 1 kohdassa tarkoitettu automatisoitu yksittäispäätös”. Toisin sanoen oli määritettävä, muodostiko ”todennäköisyysarvon automatisoitu määrittäminen... pelkästään automaattiseen käsittelyyn, kuten profilointiin, perustuvan päätöksen, jolla on rekisteröityä koskevia oikeusvaikutuksia tai joka vaikuttaa häneen vastaavalla tavalla merkittävästi” (ibid.: kohdat 20 ja 27). Vaikka SCHUFAn antama riskipisteytys välitetään kolmannelle osapuolelle, joka on viime kädessä vastuussa sopimuspäätöksen tekemisestä, tuomioistuin totesi, että kyseinen kolmas osapuoli ”käyttää tätä arvoa keskeisenä perustana” (ibid.: kohta 48). Päätöksessä esitettiin myös, että yleisen tietosuojasetuksen 22 artiklassa käytetty käsite ”päätös” ”on riittävän laaja, jotta siihen voi sisältyä henkilön luottokelpoisuuden laskennan tulos, joka perustuu todennäköisyysarvoon, joka koskee tämän henkilön kykyä täyttää maksusitoumukset tulevaisuudessa” (ibid.: kohta 46). Näin ollen, jos ”luottotietopalveluja tarjoavan yrityksen määrittämä ja pankille tiedoksi annettu todennäköisyysarvo vaikuttaa olennaisesti luoton myöntämiseen, tämän arvon määrittämisestä on pidettävä itsessään päätöksenä, jolla on rekisteröityä ’koskevia oikeusvaikutuksia tai joka vaikuttaa häneen vastaavalla tavalla merkittävästi’ yleisen tietosuojasetuksen 22 artiklan 1 kohdassa tarkoitettulla tavalla” (ibid.: kohta 50). Tämä on merkittävä päätös, joka tarjoaa oikeudellisen perustan haastaa profilointiin perustuva syrjintä ja muut päätökset, jotka perustuvat pääasiassa automaattisiin riskipisteytyksiin.

Tuomioistuin kehitti algoritmisen syrjinnän uhrien käytettävissä olevia oikeussuojakeinoja edelleen asian *Dun & Bradstreet Austria* päätöksellään (Euroopan unionin tuomioistuin, 2025). Tässä tapauksessa itävaltalainen tietoliikenneyritys kieltäytyi tekemästä hakijalle matkapuhelinsopimusta automaattisen luottokelpoisuusarvioinnin perusteella. Luottokelpoisuusarviointeihin erikoistuneen *Dun & Bradstreet* -yrityksen antaman algoritmisen riskipisteytyksen mukaan henkilöllä ei ollut riittävää luottokelpoisuutta, jotta hänelle olisi voitu myöntää kymmenen euron kuukausimaksua edellyttävä sopimus (ibid.: kohdat 2 ja 16). Kantaja otti yhteyttä Itävallan tietosuojaviranomaiseen, joka määräsi luottopisteytysyrityksen antamaan kantajalle ”merkitykselliset tiedot... henkilötietoihin perustuvaan automaattiseen päätökseen liittyvästä logiikasta” (ibid.: kohta 17). Luottokelpoisuusarviointiyritys kieltäytyi aluksi liikesalaisuuksiin vedoten antamasta hakijalle muita tietoja kuin hänen riskipistemääränsä (ibid.). Yritys oli ainoastaan todennut, että ”tämän pistemäärän saamiseksi tiettyjä [kantajan] sosiodemografisia tietoja oli ’painotettu vastaavalla tavalla’”, ja itävaltalainen tuomioistuin totesi tämän riittämättömäksi. Yrittäessään panna tämän päätöksen täytäntöön kansallinen tuomioistuin nimitti asiantuntijan, joka katsoi, että ”merkityksellisten tietojen” antamista koskevan velvollisuuden täyttäminen edellyttäisi, että *Dun & Bradstreet Austria* ilmoittaisi hakijalle, mitä henkilötietoja (esimerkiksi syntymäaika, osoite tai sukupuoli) käsiteltiin ”tekijöiden” muodostamisessa, ja seuraavat tiedot: ”kunkin kyseessä olevan tekijän osalta [kantajalle] annettu konkreettinen arvo”, ”matemaattinen kaava”, johon pistemäärä perustui, ”niiden

vaihteluvälien tarkkuus, joiden sisällä sama arvo annetaan saman tekijän eri tiedoille” sekä ”luettelo henkilöiden pistemääristä (pisteytys), jotka on saatu ajanjaksolta, joka kattaa kuusi kuukautta ennen [kantajan] pistemäärän vahvistamista ja kuusi kuukautta sen jälkeen ja joihin on päädytty saman laskentasäännön perusteella” tarkkuuden tarkistamista varten (ibid.: kohdat 23–25). Kansallisen tuomioistuimen mielestä kantajan mahdollisuus tarkistaa profiloinnin tarkkuus on ratkaisevan tärkeää täytäntöönpanon valvonnan kannalta.¹⁴ Käsiteltävänä olevassa asiassa esimerkiksi todettiin, että ”vaikka [kantajalle] toimitetut tiedot, joihin kuului muun muassa saatu pistemäärä, näyttivät, että [kantajan] luottokelpoisuus oli erittäin hyvä, tosiasiallisen profiloinnin perusteella voitiin katsoa, ettei [kantaja] ollut maksukykyinen” (ibid.: kohta 27). Merkityksellisten tietojen saanti edellyttää näin ollen kantajien mahdollisuutta haastaa syrjivä profilointi tai automaattiset päätökset yleisen tietosuojasetuksen 22 artiklan 3 kohdan mukaisesti (katso kohta 1.4.3.) (ibid.: kohta 55).¹⁵

Tuomioistuin selitti, että ”[15 artiklan 1 kohdan h alakohdan] tarkoitusten tarkastelusta ilmenee, että tässä säännöksessä tarkoitettu oikeus saada merkityksellisiä tietoja automaattiseen päätöksentekoon liittyvästä logiikasta on ymmärrettävä oikeudeksi saada selitys menettelystä ja periaatteista, joita tosiasiallisesti sovelletaan rekisteröidyn henkilötietojen hyödyntämiseen automatisoidusti tietyn tuloksen, kuten luottokelpoisuusprofiilin, saavuttamiseksi kyseisten henkilötietojen perusteella”, ja ”tämä selitys on annettava asiaankuuluvien tietojen avulla tiiviisti esitetyssä, läpinäkyvässä sekä ymmärrettävässä ja helposti saatavilla olevassa muodossa” (ibid.: kohdat 58–59). Vaikka tuomioistuin ei luettellut sovellettavia elementtejä, se totesi kuitenkin, että 15 artiklan 1 kohdan h alakohdan ”vaatimuksia ei voida täyttää pelkästään algoritmin kaltaisen monimutkaisen matemaattisen kaavan ilmoittamisella eikä automaattisen päätöksenteon kaikkien vaiheiden yksityiskohtaisella kuvauksella, koska kumpikaan näistä ei ole riittävän tiivis ja ymmärrettävä selitys” (ibid.: kohta 59). Toisin sanoen:

yleisen tietosuojasetuksen 15 artiklan 1 kohdan h alakohdassa tarkoitetuissa merkityksellisissä tiedoissa automaattiseen päätöksentekoon liittyvästä logiikasta on kuvattava menettely ja periaatteet, joita tosiasiallisesti sovelletaan, siten, että rekisteröity voi ymmärtää, mitä hänen henkilötietojaan on käytetty ja millä tavoin kyseessä olevassa automaattisessa päätöksenteossa, eikä automaattisen päätöksenteon yhteydessä toteutettavien toimien monimutkaisuus voi vapauttaa rekisterinpitäjää selitysvastuusta (ibid.: kohta 61).

Tuomiossa todettiin, että ”tuomioistuin voisi erityisesti... katsoa riittävän läpinäkyväksi ja ymmärrettäväksi muun muassa sen, että rekisteröidylle ilmoitetaan, missä määrin huomioon otettujen henkilötietojen muuttuminen olisi johtanut erilaiseen

14. Kansallinen tuomioistuin totesi, että ”siinä tapauksessa, että yleisen tietosuojasetuksen 15 artiklan 1 kohdan h alakohdassa ei taata tätä tarkistamista, oikeus saada tutustua rekisteröidyn henkilötietoihin ja muihin siinä säädettyihin tietoihin menettää merkityksensä ja tekee siitä tarpeettoman, varsinkin kun jokainen rekisterinpitäjä voisi tässä tapauksessa antaa virheellisiä tietoja” (ibid., kohta 29).

15. Tuomioistuimen mukaan ”siinä erityisessä asiayhteydessä, jossa tehdään päätös, joka perustuu yksinomaan automaattiseen käsittelyyn, rekisteröidyn yleisen tietosuojasetuksen 15 artiklan 1 kohdan h alakohdassa säädettyjen tietojen saamista koskevan oikeuden pääasiallisena tarkoituksena on antaa rekisteröidylle mahdollisuus käyttää tehokkaasti hänelle kyseisen asetuksen 22 artiklan 3 kohdassa tunnustettuja oikeuksia eli oikeutta ilmaista kantansa kyseisestä päätöksestä ja oikeutta riitauttaa se” (ibid.: kohta 55).

lopputulokseen” (ibid.). Syrjinnän vastaisen lainsäädännön kannalta tämä oikeus saada selvitys on olennainen, koska sen avulla kantajat voivat verrata itseään todellisiin tai hypoteettisiin vertailukohteisiin. Tämä on syrjintätestin olennainen osatekijä, mutta se on kuitenkin muuttunut erittäin vaikeasti saavutettavaksi algoritmisen läpinäkyvyyden ja episteemisen pirstaloitumisen vuoksi (katso kohta 1.4.3.). Jos liikesalaisuudet voivat vaarantua, ”kyseisen rekisterinpitäjän on toimitettava nämä väitetysti suojatut tiedot toimivaltaiselle valvontaviranomaiselle tai tuomioistuimelle, jonka on punnittava kyseessä olevia oikeuksia ja etuja mainitun asetuksen 15 artiklassa säädetyn rekisteröidyllä olevan tietoihin pääsyä koskevan oikeuden laajuuden määrittämiseksi” (ibid.: kohta 76).

Ranskassa 15 organisaatiota vei hiljattain korkeimman hallinto-oikeuden eli Conseil d’État’n käsiteltäväksi CNAF:n (Caisse nationale des allocations familiales, Ranskan perhe-etuuksista vastaava sosiaalivirasto) päätöksen käyttää riskipisteytysalgoritmia, jonka avulla päätettiin, keitä tuensaajia tutkittaisiin petosten osalta. ADM-järjestelmän oli osoitettu aiheuttavan muun muassa naisiin ja köyhyydessä eläviin ihmisiin kohdistuvia syrjiviä vaikutuksia. Asianomaiset pyysivät Conseil d’État’ta esittämään tuomioistuimelle kysymyksiä, jotka koskivat muun muassa algoritmista syrjintää. Tätä tapausta on seurattava, koska se antaa tuomioistuimelle mahdollisuuden tarkastella algoritmista syrjintää välittömän ja välillisen syrjinnän näkökulmasta.

1.5. Keskeiset sidosryhmät

Euroopan tekoälytoimisto

Useilla instituutioilla on rooli algoritmisen syrjinnän torjunnassa. Hiljattain perustettu Euroopan tekoälytoimisto¹⁶ (perustettu Euroopan komission yhteyteen) tukee tekoälyasetuksen täytäntöönpanoa ja tutkii mahdollisia rikkomuksia. Euroopan tekoälytoimisto vastaa 90 artiklan mukaan erityisesti riippumattomista asiantuntijoista koostuvan tiedelautakunnan (tekoälyasetuksen 68 artikla) tekemien, systeemisistä riskejä aiheuttavien GPAl-mallien varoitusten käsittelystä. Markkinavalvontaviranomaisen tavoin tekoälytoimistolla on myös valtuudet seurata ja valvoa GPAl-malleja sisältävien tekoälyjärjestelmien vaatimustenmukaisuutta (tekoälyasetuksen 75 artiklan 1 kohta). Se tekee myös yhteistyötä markkinavalvontaviranomaisten kanssa seurataksaan, tutkiakseen ja arvioidakseen suuririskisillä alueilla käytettävien GPAl-mallien vaatimustenmukaisuutta (tekoälyasetuksen 75 artiklan 2 kohta). Tekoälyasetuksen 75 artiklan 3 kohdan mukaan, jos ”markkinavalvontaviranomainen ei pysty saattamaan suuririskistä tekoälyjärjestelmää koskevaa tutkimustaan päätökseen, koska se ei voi saada tiettyjä yleiskäyttöiseen tekoälymalliin liittyviä tietoja, vaikka se on toteuttanut kaikki asianmukaiset toimet kyseisten tietojen saamiseksi, se voi esittää tekoälytoimistolle perustellun pyynnön, jonka mukaisesti pääsy kyseisiin tietoihin on pantava täytäntöön”. Näiden säännösten perusteella on todennäköistä, että tekoälyasetuksen 77 artiklan mukaisesti nimetyt yhdenvertaisuuselimet tekevät yhteistyötä tekoälytoimiston kanssa suoraan tai välillisesti markkinavalvontaviranomaisten kanssa tehtävän yhteistyön puitteissa käsitellessään syrjinnän vastaisiin oikeuksiin kohdistuvia riskejä

16. Katso <https://digital-strategy.ec.europa.eu/en/policies/ai-office>, viitattu 10. marraskuuta 2025.

tai syrjinnän vastaisten oikeuksien loukkauksia. Kuten kohdassa 1.4.1. on kuvattu, tekoölyasetuksen 73 artiklan 7 kohdassa edellytetään, että sovellettavat markkina-valvontaviranomaiset ilmoittavat tekoölyasetuksen 77 artiklan 1 kohdan mukaisille kansallisille viranomaisille tai muille elimille mahdollisista vakavista vaaratilanteista, joilla tarkoitetaan ”perusoikeuksien suojaamiseen tarkoitettujen unionin oikeuden mukaisten velvoitteiden rikkomi[sta]” tekoölyasetuksen 3 artiklan 49 kohdan c alakohdan mukaan. Tekoölyasetuksen 79 artiklan 2 kohdassa edellytetään myös, että ”jos perusoikeuksiin liittyviä riskejä todetaan”, markkina-valvontaviranomaisten on ”ilmoitettava asiasta myös 77 artiklan 1 kohdassa tarkoitetuille asiaankuuluville kansallisille viranomaisille tai elimille ja tehtävä tiivistä yhteistyötä niiden kanssa”. Myös tekoölyasetuksen 77 artiklan 1 kohdan mukaisten viranomaisten konsultointi on pakollista tekoölyasetuksen 82 artiklan 1 kohdan mukaan.

Euroopan tekoölyneuvosto

EU:n tekoölyasetuksen 65 artiklassa määrätään myös Euroopan tekoölyneuvoston perustamisesta.¹⁷ Tekoölyneuvosto koostuu jäsenvaltioiden edustajista, ja Euroopan tietosuojavaltuutettu osallistuu siihen tarkkailijana. Tekoölyneuvostolla on tärkeä rooli tekoölyasetuksen täytäntöönpanossa. Se auttaa erityisesti ”koordinoimaan EU:n jäsenvaltioiden välistä yhteistyötä ja varmistamaan tekoölyasetuksen yhdenmukaisen täytäntöönpanon ja käyttöönoton koko unionissa” (Euroopan komissio, 2025c). Se tarjoaa yhteistyöalustan tekoölyasetuksen täytäntöönpanosta vastaaville kansallisille toimivaltaisille viranomaisille ja helpottaa teknisen ja sääntelyllisen asiantuntemuksen sekä parhaiden käytäntöjen vaihtoa eri viranomaisten välillä. Sen tehtävänä on myös perustaa alaryhmiä helpottamaan yhteistyötä täytäntöönpanokysymyksissä. Tämä tarkoittaa erityisesti kahta pysyvää alaryhmää, jotka tarjoavat alustan yhteistyölle ja viestinnälle markkina-valvontaviranomaisten ja ilmoittamisesta vastaavien viranomaisten kesken ja antavat tekoölypolitiikkaa koskevia neuvoja.

Tekoölyä käsittelevä neuvoa-antava foorumi

Tekoölyasetuksen 67 artiklassa määrätään neuvoa-antavan foorumin perustamisesta¹⁸. Sen tehtävänä on tarjota teknistä asiantuntemusta ja neuvoja sekä tukea tekoölyneuvoston ja Euroopan komission työtä. Neuvoa-antavalla foorumilla tulee olla ”tasapainoinen edustus eri sidosryhmistä, joita ovat muun muassa toimiala, startup-yritykset, [pienet ja keskiuuret] pk-yritykset, kansalaisyhteiskunta ja tiedeyhteisö”. Kaupallisten ja ei-kaupallisten etujen, sekä kaupallisten etujen osalta pk-yritysten ja muiden yritysten, on oltava tasapainoisesti edustettuina jäsenistössä.

Riippumattomista asiantuntijoista koostuva tiedelautakunta

Tekoölyasetuksen 68 artiklassa määrätään riippumattomista asiantuntijoista koostuvan tiedelautakunnan perustamisesta. Sen tavoitteena on tukea asetuksen täytäntöönpanotoimia. Sen jäsenenä on riippumattomia asiantuntijoita, jotka komissio valitsee tehtävien edellyttämän tekoölyn alan viimeisimmän tieteellisen tai teknisen

17. Katso <https://artificialintelligenceact.eu/article/65>, viitattu 10. marraskuuta 2025.

18. Katso <https://artificialintelligenceact.eu/article/67/>, viitattu 10. marraskuuta 2025.

asiantuntemuksen perusteella ja joiden on kyettävä osoittamaan täyttävänsä kaikki seuraavien tehtävien edellytykset:

- ▶ tukea tekoälyasetuksen täytäntöönpanoa ja noudattamisen valvontaa erityisesti GPAI-mallien osalta (esimerkiksi varoittamalla tekoälytoimistoa mahdollisista systeemisistä riskeistä, osallistumalla sellaisten välineiden ja menetelmien kehittämiseen, joilla arvioidaan GPAI-mallien ja -järjestelmien valmiuksia, antamalla neuvoja erilaisten yleiskäyttöisten tekoälymallien ja -järjestelmien luokittelusta riippumatta siitä, liittyykö niihin systeemisiä riskejä vai ei, ja osallistumalla välineiden ja mallien kehittämiseen)
- ▶ tukea markkinavalvontaviranomaisten työtä niiden pyynnöstä
- ▶ tukea rajat ylittäviä markkinavalvontatoimia
- ▶ tukea tekoälytoimistoa sen suorittaessa unionin suojamenettelyn mukaisia tehtäviään.

CEN-CENELEC

CEN-CENELECin tekninen sekakomitea 21 (Joint Technical Committee, JTC 21) työskentelee tekoälyä koskevien standardien parissa Euroopan komission standardointipyyntöä.¹⁹ JTC 21 laatii parhaillaan yhdenmukaistettua standardia, joka koskee tekoälyasetuksen 9 artiklassa kuvattua suuririskisten tekoälyjärjestelmien tarjoajien vaatimusta käyttää riskinhallintajärjestelmää.

Yhdenvertaisuuselinten rooli

Yhdenvertaisuuselimillä on keskeinen rooli syrjinnän torjunnassa. Algoritmista syrjintää tapahtuu usein tavoilla, joita on vaikea havaita ja todistaa. Tämä muodostaa merkittävän esteen asianosaisten oikeussuojan saatavuudelle. Algoritmisen syrjinnän uhrien voi olla vaikea ymmärtää, miten tai miksi heitä kohdeltiin epäoikeudenmukaisesti, mikä vaikeuttaa näiden päätösten riitauttamista. Yhdenvertaisuuselimillä on näin ollen ratkaisevan tärkeä rooli uhrien tukemisessa ja vastuuvollisuuden edistämisessä. Tässä kontekstissa hiljattain hyväksytyillä yhdenvertaisuuselimiä koskevilla direktiiveillä (direktiivi 2024/1500 ja direktiivi 2024/1499) vahvistetaan yhdenvertaisuuselinten riippumattomuutta, resursseja ja oikeussuojan saatavuuden helpottamista koskevaa roolia sekä niitä koskevia valtuuksia. Niillä vahvistetaan myös yhdenvertaisuuselinten rooli tekoälyyn perustuvan syrjinnän käsittelyssä:

On tärkeää kiinnittää huomiota automatisoitujen järjestelmien, kuten tekoälyn, käytön tarjoamiin mahdollisuuksiin ja riskeihin. Tasa-arvoelimillä olisi erityisesti oltava käytettävissään sopivat henkilöstö- ja tekniset resurssit. Kyseisten resurssien olisi etenkin mahdollistettava se, että tasa-arvoelimet voivat käyttää automatisoituja järjestelmiä työssään ja toisaalta arvioida sitä, ovatko kyseiset järjestelmät syrjimättömyyssääntöjen mukaisia. Jos tasa-arvoelin on osa elintä, jolla on useita tehtäviä, sen tasa-arvotehtävän hoitamiseen tarvittavat resurssit olisi varmistettava (Euroopan unioni, 2024b: johdanto-osan kappale 21; Euroopan unioni, 2024a: johdanto-osan kappale 22).

19. Euroopan komission täytäntöönpanopäätös, annettu 22 päivänä toukokuuta 2023, sisältäen standardointipyynnön C(2023)3215.

Tässä yhteydessä yhdenvertaisuuselimillä voi olla kriittisen tärkeä rooli uhrien tukemisessa, tietoisuuden lisäämisessä, tutkimuksen tekemisessä sekä tekoälyn ja ADM-järjestelmien paremman sääntelyn edistämisessä:

- ▶ Viran puolesta toimiminen ja algoritmisen syrjinnän uhrien valitusten käsittely: Yhdenvertaisuuselimillä voi olla viran puolesta tutkintavaltuudet, ja ne toimivat helposti saatavilla olevana yhteyspisteenä algoritmisen syrjinnän mahdollisille uhreille. Ne voivat käsitellä uhrien tekemiä valituksia, tutkia väitettyjä algoritmisten vinoumien tapauksia ja antaa neuvoja ja tukea oikeustoimissa.
- ▶ Oikeudellisen avun antaminen ja tapausten vieminen tuomioistuimeen: tähän voi kuulua algoritmiseen syrjintään liittyvien tapausten ratkaiseminen, yksilöiden neuvominen siinä, miten he voivat riitauttaa tekoälyn tai ADM-järjestelmien tekemät syrjivät päätökset, tai joissakin tapauksissa mahdollisten uhrien edustaminen tai heidän puolestaan toimiminen tuomioistuimissa.
- ▶ Tietoisuuden lisääminen ja tietojen antaminen uhreille ja suurelle yleisölle: yhdenvertaisuuselimillä voi olla keskeinen rooli yleisön valistamisessa algoritmista syrjinnästä mahdollisesti aiheutuvista haitoista esimerkiksi antamalla tietoa siitä, miten epäoikeudenmukaiset automaattiset päätökset voidaan haastaa ja mitä tekoäly- ja ADM-prosesseihin liittyviä oikeuksia yksilöillä on.
- ▶ Tietojen kerääminen ja tekoäly- ja ADM-järjestelmiin liittyvää syrjintää koskevien tutkimusten tekeminen: Toinen yhdenvertaisuuselinten kriittinen tehtävä on kerätä tietoja algoritmista syrjinnästä. Seuraamalla järjestelmällisesti tekoälyyn ja ADM-järjestelmiin liittyviä valituksia ja tekemällä tutkimusta, tiedonkeruuta ja seuranta yhdenvertaisuuselimet voivat tunnistaa syrjintämalleja ja tuoda esiin systeemisii ongelmia, jotka saattavat edellyttää sääntelytoimia.
- ▶ Suositusten antaminen poliittisille päättäjille ja lainsäätäjille: Tällaiset tiedot voivat myös antaa konkreettista näyttöä lainsäädännön uudistusten tarpeesta (esimerkiksi rekisteröinti- ja tiedotusvelvollisuuksien osalta). Yhdenvertaisuuselimet voivat antaa valvontatoimien perusteella näyttöön perustuvia suosituksia poliittisille päättäjille ja lainsäätäjille siitä, miten tekoäly ja ADM-järjestelmiä voitaisiin säännellä, erityisesti täytäntöönpanon valvonnan ja siihen liittyvien haasteiden näkökulmasta.
- ▶ Sidosryhmien kanssa tehtävä yhteistyö yhdenvertaisuuden edistämiseksi tekoäly- ja ADM-järjestelmissä: Yhdenvertaisuuselimet voivat tehdä yhteistyötä monien eri sidosryhmien, kuten työnantajien, palveluntarjoajien, julkisten laitosten ja kansalaisjärjestöjen, kanssa edistääkseen hyvien tekoäly- ja ADM-järjestelmiin liittyvien yhdenvertaisuuskäytäntöjen käyttöönottoa. Tämä voi tarkoittaa esimerkiksi sitä, että organisaatioita autetaan kehittämään ja toteuttamaan yhdenvertaisuussuunnitelmia, joilla minimoidaan algoritmisten vinoumien riski, ja annetaan ohjeita siitä, miten tekoälyn ja ADM-järjestelmien mahdollista syrjintää voidaan testata.



2. Oikeudelliset puutteet

Tässä osiossa tunnistetaan nykyisten oikeudellisten, menettelyllisten ja hallinnollisten puutteita ja tarpeita, kun otetaan huomioon tekoälyn ja ADM-järjestelmien käytöstä julkishallinnossa ja muilla tämän tutkimuksen piiriin kuuluvilla sektoreilla aiheutuvat syrjintäriskit. Osiossa esitetään myös yhteenveto tärkeimmistä esteistä, puutteista ja tarpeista, joita yhdenvertaisuuselimet ja muut asiaankuuluvat sidosryhmät kohtaavat seuranta- ja valvontatehtävissään.

2.1. Perustelut tekoälyn ja ADM-järjestelmien käyttöönotolle

Osiossa I on osoitettu, että tekoäly- ja ADM-teknologioiden käyttöönotto yleistyy jatkuvasti. Niiden käyttöön liittyy kuitenkin useita ongelmia. Ensinnäkin näiden järjestelmien käyttöönottoon kannustavat narratiivit ovat ongelmallisia ja lisäävät oikeudellisia riskejä. Tekoäly- ja ADM-järjestelmät esitetään usein "neutraaleina" tai ainakin "vähemmän puolueellisia" ihmisten tekemiin päätöksiin verrattuina. Niiden käyttöä saatetaan perustella halulla rationalisoida päätöksentekoprosesseja. Koska näiden järjestelmien tiedetään sisältävän vinoumia laajalti, tällainen narratiivi voi luoda illuusion siitä, että tekoälyn ja ADM-järjestelmien käyttöönotto vähentää syrjintää tai jopa estää sen. Tämä aiheuttaa riskin siitä, että syrjinnän vastaisen lainsäädännön rikkomukset jäävät huomiotta.

Lisäksi tekoälyä ja ADM-järjestelmiä mainostetaan usein keinona parantaa päätöksentekoprosessien kustannustehokkuutta etenkin julkishallinnossa, jota tarkkaillaan tarkasti menoleikkausten ja talouden kiristämisen konteksteissa. Tällaiset perustelut johtavat kulutuksen optimoijina esitettävien tekoäly- ja ADM-järjestelmien käyttöönottoon tietyissä käyttötarkoituksissa, kuten petosten havaitsemisessa tai riskin-arvioinnissa. Tällainen narratiivi peittää kuitenkin ainakin kaksi tärkeää kysymystä:

- ▶ Pitäisikö tällaisia teknologioita ottaa käyttöön vaihtoehtoisin käyttötarkoituksiin, esimerkiksi oikeuksien ja julkisten palvelujen saatavuuden parantamiseksi, sillä tutkimukset osoittavat säännöllisesti, ettei niitä ole riittävästi hyödynnetty tällaisissa kohteissa? (Défenseur des Droits, 2024).
- ▶ Voidaanko tällaisten järjestelmien avulla todella vähentää julkisia menoja ja säilyttää samalla julkisten palvelujen laatu, kun kaikki perusoikeuksien turvaamiseen liittyvät kustannukset on otettu huomioon?

On esimerkiksi kallista estää syrjintäriskien toteutuminen riittävillä testaus-, seuranta- ja tarkastusmenettelyillä, kouluttaa algoritmisia suosituksia käsitteleviä työntekijöitä tehokkaan ihmisten tekemän valvonnan varmistamiseksi sekä luoda tehokkaita tiedotus- ja muutoksenhakumekanismeja julkisten palvelujen käyttäjille. Julkishallinnon on todennäköisesti huomioitava tällaiset kustannukset, kun se tekee päätöksiä näiden järjestelmien käyttöönotosta. Tällaisen harkinnan tulisi myös saada julkiset päätöksentekijät esittämään ”nollakysymyksiä”: jos tekoäly- tai ADM-järjestelmän käyttöönotto vaikuttaa liian riskialttiilta tai kalliilta sen jälkeen, kun ennaltaehkäisevät ja perusoikeuksia turvaavat toimet on toteutettu asianmukaisesti, olisi päätettävä, ettei kyseistä järjestelmää oteta käyttöön.

2.2. Avoimuuden puute

Kuten jäljempänä olevissa soveltuissa kohdissa todetaan, tekoäly- ja ADM-järjestelmien kehittämisestä, testauksesta ja käyttöönotosta julkishallinnossa ja sen ulkopuolella ei yleisesti ottaen ole systemaattista, selkeää ja helposti saatavilla olevaa tietoa. Tekoälyn ja ADM-järjestelmien käyttöä julkishallinnossa ei kartoiteta systemaattisesti, vaikka tällä hetkellä on suunnitelmia tai pyrkimyksiä luoda tietokantoja tällaisesta käytöstä (esimerkiksi Alankomaiden algoritmirekisteri).²⁰ Algoritmista syrjintää koskevia tapauksia on tällä hetkellä vielä vähän, ja ne liittyvät usein enemmänkin tietosuojaan kuin syrjintään ja yhdenvertaisuuteen. Näitä harvoja tapauksia on usein käsitelty tuomioistuimissa yhdenvertaisuuselinten ja kansalaisjärjestöjen tuella. Tietojen puute siitä, milloin, miksi ja missä julkishallinnot käyttävät tekoäly- ja ADM-järjestelmiä – ja näin ollen toistaiseksi sovellettavien oikeudellisten raportointivelvoitteiden puuttuminen – vaikeuttaa kuitenkin näiden toimijoiden tehokasta ja systemaattista algoritmisen syrjinnän seurantaa, arviointia ja haastamista tutkintavaltuuksista huolimatta. Jotta nämä toimijat voisivat käyttää tutkintavaltuuksiaan tehokkaasti, heidän on ennen kaikkea oltava tietoisia tekoäly- tai ADM-sovellusten olemassaolosta. Yhdenvertaisuuselinten on kuitenkin usein turvauduttava median tekemiin tutkimuksiin, alaan erikoistuneisiin kansalaisjärjestöihin tai epävirallisiin verkostoihin, jotta ne voivat tunnistaa mahdollisia algoritmisen syrjinnän tapauksia.

20. Katso <https://algoritmes.overheid.nl/en>, viitattu 10. marraskuuta 2025.

Vaikka kansalliset asiantuntijat raportoivat, että viranomaisille lähetetään säännöllisesti tietopyyntöjä (esimerkiksi tiedonvälityksen vapauden nojalla), jotta tiedot tekoälyn ja ADM-järjestelmien käytöstä julkishallinnossa saataisiin julkiseksi, pyyntöjen oikeellisuus kyseenalaistetaan toisinaan tai niihin ei vastata, jos ne eivät ole sitovia ja erityisesti silloin, kun pyynnöt koskevat arkaluonteisia sovelluksia, kuten digitaalista valvontateknologiaa. Lisäksi tietojen pyytäjät eivät usein pysty tarkistamaan saamiaan tietoja. Toisinaan käytetään myös muita tiedonhankintakanavia, kuten hallitukselle yhteistyössä kansallisten parlamenttien jäsenten kanssa esitettyjä kysymyksiä. Sekä tiedonvälityksen vapautta koskevien pyyntöjen että parlamentissa esitettyjen kysymysten tehokkuutta heikentää kuitenkin tekoäly- ja ADM-sovellusten käytön avoimuuden puute. Tiedot yhdenvertaisuuselimet tarjoavat myönteisiä esimerkkejä tästä: esimerkiksi Suomen yhdenvertaisuusvaltuutettu voi lähettää julkishallinnolle pakollisia tieto- ja selvityspyyntöjä. Tällaiset valtuudet voivat helpottaa tekoälyn ja ADM-järjestelmien käyttöön ja niiden mahdollisiin syrjiviin vaikutuksiin liittyvää tiedonsaantia ja avoimuutta.

Tietojen puutteeseen puututaan jossain määrin uusilla tekoälyä koskevilla asetuksilla. Esimerkiksi EU:n tekoälytietokantaan, josta säädetään tekoälyasetuksen 71 artiklassa, sisällytetään julkisia tietoja suuririskisistä tekoälysovelluksista.²¹ Lisäksi tekoälyä koskevan puiteyleissopimuksen 14 artiklan a–b kohdassa vaaditaan, että ”asiaankuuluvat tiedot tekoälyjärjestelmistä, jotka voivat vaikuttaa merkittävästi ihmisoikeuksiin, ja niiden merkityksellisestä käytöstä dokumentoidaan, toimitetaan elimille, joilla on valtuudet tutustua kyseisiin tietoihin, ja tarvittaessa ja soveltuvin osin asetetaan asianomaisten henkilöiden saataville tai annetaan heille tiedoksi”, jotta nämä henkilöt voivat ”riitauttaa päätökset, jotka on tehty järjestelmää käyttäen tai joihin järjestelmän käyttö on vaikuttanut merkittävästi”. Näitä määräyksiä vahvistetaan edelleen puiteyleissopimuksen avoimuutta ja valvontaa sekä vastuuvälillisyyttä ja vastuuta koskevalla 8 ja 9 artiklalla. Tällaiset tiedotus- ja avoimuusvaatimukset voivat mahdollistaa tietojen julkisen tarkastelun. Vastaavasti puiteyleissopimuksen 16 artiklassa edellytetään riskin- ja vaikutustenhallintakehityksen käyttöönottoa ja voimassa pitämistä (tätä tukee HUDERIA-menetelmän standardoitu vaikutustenarviointi), ja 16 artiklan 2 kohdan f alakohdassa todetaan, että ”niihin on sisällyttävä riskien, todellisten ja mahdollisten vaikutusten sekä riskinhallintaan sovellettavan lähestymistavan dokumentointi”. Tällaisten tietojen saatavuus voi auttaa perusoikeuksien suojelusta vastaavia elimiä ehkäisemään algoritmista syrjintää ja puuttumaan siihen. Tekoälyasetuksessa puolestaan säädetään, että 77 artiklan mukaisesti nimeillä viranomaisilla on valtuudet saada käyttöönsä tekoälyasetukseen perustuva tekninen dokumentaatio ja riskinarviointijärjestelmät sekä perusoikeusvaikutusten arvioinnit, kun ne ovat pakollisia tekoälyasetuksen 27 artiklan nojalla (rajoitettu tiettyihin käyttöönottajiin, jotka tarjoavat julkisia tai olennaisia yksityisiä palveluita).

Tällaiset tiedotus- ja avoimuusvelvoitteet olisi kuitenkin voitu muotoilla laajemmin. Tällä hetkellä vaatimukset koskevat vain suuririskiä tekoälyjärjestelmiä eli ulkopuolelle jäävät järjestelmät, joita ei ole kuvattu suuririskisiksi järjestelmiksi, sekä

21. Lainvalvonnan, muuttoliikkeen hallinnan, turvapaikka-asoiden ja rajavalvonnan alan suuririskisiä tekoälyjärjestelmiä koskevat tiedot, mukaan lukien biometriikkaa koskevat tiedot, on rekisteröitävä tietokannan ei-julkiseen osioon. Pääsy ei-julkiseen osioon on rajattu Euroopan komissiolle sekä erityisille kansallisille markkinavalvontaviranomaisille (esimerkiksi tietosuojaviranomaisille).

ADM-järjestelmät, joiden ei katsota kuuluvan tekoälyn määritelmän piiriin. Vuonna 2020 Euroopan neuvoston parlamentaarinen yleiskokous vaati, että hallituksia pyydetään ”ilmoittamaan parlamentille ennen [tekoäly- ja ADM-järjestelmien] teknologian käyttöönottoa” ja että ”tällaisen teknologian viranomaiskäyttö... kirjataan järjestelmällisesti julkiseen rekisteriin” (Euroopan neuvosto, parlamentaarinen yleiskokous, 2020b). Tämä ei näy suoraan äskettäin hyväksytyssä tekoälyä koskevassa puiteyleis-sopimuksessa, jossa ei vaadita julkisen rekisterin perustamista. Tekoälyasetuksessa ei puolestaan säädetä ilmoituksista kansallisille parlamenteille tai EU:n toimielimille, kun viranomaiset ottavat käyttöön tekoäly- tai ADM-järjestelmiä.

2.3. Oikeussuojan saatavuushaasteet

Oikeussuojan saatavuutta vaikeuttavat erilaiset tiedonsaannin ja vallan epäsymmetriat. Kuten edellä selitettiin, yksi kansallisella tasolla tehdyn tutkimuksen tärkeimmistä havainnoista osoittaa, että tekoälyn ja ADM-järjestelmien käytön selkeän ja johdonmukaisen kartoituksen puute julkishallinnossa vaarantaa oikeuden syrjimättömyyteen ja sen soveltamiseen. Kun ei tiedetä, että tietty hallinnollinen päätös on tehty tekoälyllä tai ADM-järjestelmällä (tai sen tuella), tällaisten päätösten kohteena olevat henkilöt eivät voi edes epäillä algoritmista syrjintää. Vaikka epäilyjä heräisi, julkishallinnolle ei ole asetettu yleistä velvollisuutta antaa tietoja käytetyn järjestelmän tyypistä, päätöksenteossa sovelletuista tekijöistä ja tavasta, jolla niitä käytetään päätöksentekoprosessissa. Tämä estää uhreja esittämästä tuomioistuimissa riittäviä todisteita, joilla voitaisiin osoittaa syrjintäoletta. ²²

EU:n syrjinnän vastaisessa lainsäädännössä säädetään toki todistustaakan jakamisesta kantajien ja vastaajien kesken, jotta kantajien taakka syrjintää koskevissa tapauksissa helpottuisi. EU:n syrjinnän vastaisen lainsäädännön mukaan, ”kun henkilö, joka katsoo kärsineensä vääryyttä sen vuoksi, että häneen ei ole sovellettu yhdenvertaisen kohtelun periaatetta, esittää tuomioistuimessa tai muussa toimivaltaisessa viranomaisessa tosiseikkoja, joiden perusteella voidaan olettaa tapahtuneen välitöntä tai välillistä syrjintää, vastaajan on näytettävä toteen, ett[ä] yhdenvertaisen kohtelun periaatetta ei ole rikottu”.²³ Vastaavasti Euroopan ihmisoikeustuomioistuin siirtää todistustaakan syrjimättömyyden osoittamisesta vastaajille, kun kantaja esittää luotettavia ja johdonmukaisia tosiseikkoja, joiden perusteella voidaan olettaa, että syrjintää on tapahtunut: ”todisteet voivat seurata samanaikaisia, riittävän vahvoja, selkeitä ja yhtäpitäviä päättelyitä tai vastaavia kiistattomia tosiseikkojen olettamia” (Euroopan ihmisoikeustuomioistuin, 2007). Periaatteessa nämä säännöt helpottavat kantajien tehtävää edellyttämällä, että heidän on vain osoitettava syrjintäoletta, ennen kuin todistustaakka voidaan siirtää vastaajalle. Tämän jälkeen vastaajan tehtävänä on kumota syrjintäoletta.

Näissä säännöissä ei kuitenkaan käsitellä riittävästi vallan ja tiedon epäsymmetrioita, joita ilmenee, kun päätös tehdään tekoälyllä tai ADM-järjestelmällä (tai niiden tuella),

22. Tätä ongelmaa voidaan jossain määrin lieventää käyttämällä yleisessä tietosuoja-asetuksessa säädettyä oikeutta saada selvitys, vaikka menettelyyn liittyvät esteet ja kustannukset voivat olla yksittäisille uhreille suuria.

23. Katso direktiivin 2000/43/EY 8 artiklan 1 kohta. Tämä säännös sisältyy kaikkiin syrjinnän vastaisiin direktiiveihin: 2000/78/EY, 2004/113/EY ja 2006/54/EY.

ja ne voivat olla riittämättömiä algoritmisen syrjinnän tapauksissa. Ensinnäkin kansalliset asiantuntijat ovat raportoineet, että tuomioistuimet asettavat toisinaan korkean todistuskynnyksen todistustaakan siirtämiselle vastaajille. Toisaalta algoritmisen läpinäkyvyyden puute vaikeuttaa yksittäisten kantajien tehtävää, kun he pyrkivät osoittamaan oletettaman algoritmisen syrjinnän (katso aiempi kohta 2.2.). Tiedon ja vallan epäsymmetriaa ilmenee myös tavalla, jota Milano ja Prunkl (2024) kutsuvat ”episteemiseksi pirstaloitumiseksi”. Sillä tarkoitetaan ”yksilöitä eristävää algoritmivälitteisten ympäristöjen rakenteellista ominaisuutta”, joka ”estää ihmisiä ja yhteisöjä jakamasta ja vertailemasta kokemuksiaan merkityksellisesti”. Episteeminen pirstaloituminen näin ollen ”vaikeuttaa... näissä ympäristöissä ilmenevien haittojen tunnistamista ja käsitteellistämistä”. Tähän vaikuttaa erityisesti se, että eurooppalainen yhdenvertaisuuslainsäädäntö perustuu uhrien kykyyn osoittaa, että heitä on kohdeltu eriarvoisesti tai heidät on saatettu epäsuotuisen asemaan verrattuna muihin samankaltaisessa tilanteessa oleviin henkilöihin (ibid.). Algoritmisen läpinäkyvyyden puute ja episteeminen pirstaloituminen yhdessä tekevät syrjivien tekoäly- ja ADM-järjestelmien oikeudellisesta haastamisesta hyvin vaikeaa.

Näin ollen oikeussuojan saatavuus vaarantuu käytännössä. Konkreettisenä esimerkkinä mainittakoon, että useiden eurooppalaisten sosiaaliturvaviranomaisten petosten havaitsemiseen käyttämät ADM-järjestelmät luovat ”epäilyttävien henkilöiden” *de facto* -luokan, joka on erittäin syrjivä, mutta jota kyseisten henkilöiden on hyvin vaikea haastaa, vaikka luokituksella voi olla vakavia seurauksia heidän elämilleen (sosiaaliturvaetuuksien lakkauttaminen, perusteettomasti maksettujen tukien palautuspyynnöt ja niin edelleen). Vaikka unionin tuomioistuimen hiljattain vahvistama oikeus saada selvitys yleisen tietosuojasetuksen 15 artiklan 1 kohdan h alakohdan ja 22 artiklan mukaisesti Dun & Bradstreet Austria -asiassa osittain ratkaisee tämän ongelman (katso kohta 1.4.4.), vallan ja tiedonsaannin epäsymmetrioita on edelleen olemassa, ja niihin on puututtava keventämällä kantajien todistustaakkaa entisestään antamalla yhdenvertaisuuselimille ja asiaankuuluville sidosryhmille valtuudet tukea mahdollisia uhreja ja lisäämällä ennaltaehkäisytoimia.²⁴ Euroopan neuvoston parlamentaarinen yleiskokous toi tämän ongelman esille jo vuonna 2020 ja korosti tarvetta ”kiinnittää erityistä huomiota syyttömyysolettaman takaamiseen ja sen varmistamiseen, että syrjinnän uhreille ei aseteta suhteetonta todistustaakkaa” (Euroopan neuvosto, parlamentaarinen yleiskokous, 2020b). Lisäksi jäsenvaltioita kehoitettiin ”varmistamaan, että yhdenvertaisuuselimillä on täydet valtuudet käsitellä tekoälyn käytöstä johtuvia yhdenvertaisuuteen ja syrjimättömyyteen liittyviä kysymyksiä”.

Tästä näkökulmasta joidenkin eurooppalaisten yhdenvertaisuuselinten valtuudet edistää syrjinnän vastaista suojelua kollektiivisella ja rakenteellisella tasolla voivat auttaa lieventämään näitä puutteita. Tietyt yhdenvertaisuuselimet voivatkin osaltaan ehkäistä algoritmista syrjintää valvonta-, tutkinta- ja sanktiovaltuuksien avulla. Lisäksi jotkin yhdenvertaisuuselimet voivat valvoa tekoäly- ja ADM-järjestelmien käyttöä ja haastaa syrjivät tuotokset tuomioistuimissa yleisen edun nimissä ilman, että yksittäistä uhria tarvitsee tunnistaa etukäteen (katso Euroopan unionin tuomioistuin,

24. Tämä on nyt entistäkin tärkeämpää, koska tekoälyn liittyvää vastuuta koskeva direktiivi on peruutettu. Direktiivissä ehdotettiin, että oletettavien osoittamista ja todistustaakan siirtämistä helpotettaisiin, vaikkakin rajoitetusti.

2008). Vaikka tämä ei koske kaikkia eurooppalaisia yhdenvertaisuuselimiiä, tällaiset valtuudet voivat lieventää vaikeuksia, joita aiheutuu siitä, että yksittäisten uhrien on ilmoitettava syrjinnästä aiheutuneista haitoista, ennen kuin tutkimukset voidaan aloittaa. Jotta tällaiset valtuudet olisivat tehokkaita, niitä on kuitenkin tuettava vahvoilla tiedonsaantioikeuksilla: yhdenvertaisuuselimillä on oltava oikeus pyytää ja saada soveltuvia ja merkityksellisiä tietoja tekoöly- ja ADM-järjestelmien käyttötavoista. Tällaiset valtuudet on ulotettava koskemaan tilanteita, joissa tekoöly- ja ADM-järjestelmiä käytetään päätöksenteon tukena ja apuna, sen sijaan, että ne rajoitettaisiin koskemaan vain täysin automaattisia tilanteita. Näin voidaan ottaa huomioon se, että hybridipäätösmenettely on järjestelmien yleisimmin käytetty muoto. Yhdenvertaisuuselinten standardeja koskevien uusien direktiivien saattaminen osaksi kansallista lainsäädäntöä kesäkuuhun 2026 mennessä tarjoaa hyvän tilaisuuden tarkistaa yhdenvertaisuuselinten toimivaltuuksia näiden uusien haasteiden valossa. Lisäksi vaihtoehtoiset menettelytavat, kuten kollektiiviset valitusmekanismit (esimerkiksi *actio popularis*, ryhmäkanteet), voivat lieventää oikeussuojan saatavuuteen liittyviä ongelmia.

2.4. Nykyisten ja tulevien säännösten täytäntöönpano

2.4.1. Instituutioiden välinen yhteistyö

Tekoölyasetuksen täytäntöönpanojärjestelmä on monimutkainen eikä jätä juuriakaan tilaa yksittäisille valituksille. Vaikka markkinaevalvontaviranomaiset voivat vastaanottaa valituksia, niillä ei ole velvollisuutta tutkia asiaa niiden perusteella. Tästä näkökulmasta yhdenvertaisuuselinten rooli korostuu entisestään. Tekoölyasetuksen täytäntöönpanojärjestelmään kuuluu monia sidosryhmiä sisällyttävä yhteistyö yhdenvertaisuuselinten, muiden 77 artiklan mukaisesti nimettyjen perusoikeuksia suojaavien viranomaisten, markkinaevalvontaviranomaisten ja muiden asiaankuuluvien instituutioiden välillä. Lisäksi yhdenvertaisuuselimet tekevät yhteistyötä muiden tekoölyasetuksen 77 artiklan mukaisten perusoikeusinstituutioiden kanssa. Joidenkin tätä raporttia varten haastateltujen instituutioiden mukaan tähän saattaa kuitenkin liittyä merkittäviä haasteita. Erityisen tärkeää on kehittää ja ottaa käyttöön yhteistyömenetelmiä ja varmistaa tarvittavat institutionaaliset viestintäkanavat. Näin voidaan edistää lähestymistapaa, joka perustuu siihen, että instituutiot täydentävät toisiaan ja toimivat intersektionaalisesti. Lisäksi toimivaltaisilla viranomaisilla, mukaan lukien keskitetyillä yhteyspisteillä ja markkinaevalvontaviranomaisilla, tulee olla riittävä käsitys algoritmisten teknologioiden aiheuttamista syrjintäriskeistä sekä oikeudellisista ja institutionaalisista keinoista puuttua syrjintään.

2.4.2. Käsitteelliset puutteet

Muita täytäntöönpanoon liittyviä ongelmia syntyy eurooppalaisen syrjinnän vastaisen lainsäädännön käsitteisiin ja menettelyihin liittyen. Käytännössä tekoölyn ja erityisesti koneoppimisen toiminta perustuu korrelaatioihin, ei kausaaliiteettiin. Tämä lisää niin sanotun sijaismuuttujaan liittyvän syrjinnän riskiä. Tuomarit saattavat olla haluttomia käsittelemään tällaista syrjintää välittömänä syrjintänä, vaikka sijaismuuttujaan

liittyvän syrjinnän tiedetään korreloivan syrjintäperusteiden kanssa.²⁵ Esimerkiksi tietyissä yhteyksissä henkilöiden asuinpaikan tiedetään korreloivan voimakkaasti rodun ja etnisen alkuperän kanssa. Euroopan tasolla asuinpaikka ei sisälly syrjinnän vastaisessa lainsäädännössä kiellettyjen syrjintäperusteiden luetteloon. Näin ollen tällaista algoritmista sijaismuuttujaan liittyvää syrjintää käsitellään todennäköisesti välillisenä syrjintänä, ellei kyseessä ole tapaus, jossa asuinpaikan katsotaan olevan "erottamattomasti" tai "luonnostaan" sidoksissa kiellettyyn syrjintäperusteeseen, kuten rotuun tai etniseen alkuperään.

Tältä osin EU:n syrjinnän kieltävä sekundaarilainsäädäntö saattaa osoittautua liian rajoitetuksi, koska se koskee vain "suljettua" syrjintäperusteluetteloa: sukupuoli, rotu tai etninen alkuperä, vammaisuus, seksuaalinen suuntautuminen, uskonto tai vakaus ja ikä. Muihin sosiaalisiin ominaisuuksiin liittyviä algoritmisia haittoja pidetään välittömästi syrjivinä vain, jos niiden katsotaan "luontaisesti liittyvän" johonkin näistä kielletyistä syrjintäperusteista. Tällaisia haittoja voidaan kuitenkin pitää välillisenä syrjintänä, jos ne kohdistuvat kiellettyjä syrjintäperusteita omaaviin ryhmiin. Muut algoritmiset haitat eivät kuulu oikeussuojan piiriin, jos ne eivät sisälly erityisiin kansallisiin lainsäädäntökehyksiin, jotka voivat olla EU:n lainsäädäntöä laajempia.

Sen sijaan niin sanotut "avoimet" luettelot, kuten Euroopan unionin perusoikeuskirjan 21 artiklan 1 kohdassa ja Euroopan ihmisoikeussopimuksen 14 artiklassa kuvatut kielletyt syrjintäperusteet, voisivat tehdä algoritmisen sijaismuuttujaan liittyvän syrjinnän torjumisesta joustavampaa. Tämä johtuu siitä, että niiden avulla tuomarit voivat todeta "kaikenlaisen syrjinnän", kuten "sukupuoleen, rotuun, ihonväriin tai etniseen taikka yhteiskunnalliseen alkuperään, geneettisiin ominaisuuksiin, kieleen, uskontoon tai vakaumukseen, poliittisiin tai muihin mielipiteisiin, kansalliseen vähemmistöön kuulumiseen, varallisuuteen, syntyperään, vammaisuuteen, ikään tai sukupuoliseen suuntautumiseen tai muuhun sellaiseen seikkaan" ja "sukupuoleen, rotuun, ihonväriin, kieleen, uskontoon, poliittisiin tai muihin mielipiteisiin, kansalliseen tai yhteiskunnalliseen alkuperään, kansalliseen vähemmistöön kuulumiseen, varallisuuteen, syntyperään tai muuhun asemaan" perustuvan syrjinnän.

Kysymys siitä, miten algoritmisen sijaismuuttujaan liittyvä syrjintä määritellään, on tärkeä. Vaikka tämä määrittely ei yleissopimuksen syrjinnän vastaisen lainsäädännön mukaisesti tee vaikutusta, koska yhtenäinen järjestelmä sisältää avoimia perusteluja sekä välittömälle että välilliselle syrjinnälle, EU:n syrjinnän vastaisessa lainsäädännössä ongelma on erilainen. Siinä välitön syrjintä hyötyy tiukemmasta perustelujärjestelmästä, joka pohjautuu poikkeusten suljettuun luetteloon. Kun otetaan huomioon tekoäly- ja ADM-järjestelmien avoimuuden puute, läpinäkyväisyys ja monimutkaisuus, avoin perustelujärjestelmä, joka edellyttää kontekstuaalista analyysia, ei välttämättä ole kaikkein tarkoituksenmukaisin. Välittömän syrjinnän näyttäminen toteen on kuitenkin usein ongelmallista algoritmisessa kontekstissa.

25. Jos haluat lisätietoja sijaismuuttujaan liittyvästä syrjinnästä, katso Bartoletti I. ja Xenidis R. (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", saatavilla osoitteessa <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>, viitattu 10. marraskuuta 2025, Euroopan neuvoston julkaisut, Strasbourg.

Kuten toisaalla on esitetty, perusteluongelmien lisäksi välittömän ja välillisen syrjinnän käsitteet itsessään eivät täysin sovellu algoritmisen syrjinnän kuvaamiseen, sillä siinä sekoittuvat molempien käsitteiden piirteet (Xenidis, 2021).

Tekoäly- ja ADM-järjestelmien korkea jaottelutaso lisää myös intersektionaalisen syrjinnän riskiä. Tämä koskee erityisesti tilanteita, joissa käyttäjät yhdistävät useita erilaisia riskipisteisiin vaikuttavia riskiluokkia toisiinsa (esimerkiksi sukupuoli ja maahanmuuttohistoria vinoutuneiden työttömyysennusteiden esimerkeissä). Intersektionaalista syrjintää ei kuitenkaan vielä tunnusteta täysin EU:n syrjinnän vastaisissa direktiiveissä (lukuun ottamatta äskettäistä samapalkkaisuuden periaatteen soveltamista koskevaa direktiiviä 2023/970), eikä sitä ole viimeaikaisesta kehityksestä huolimatta kielletty yhtenäisesti kansallisessa lainsäädännössä. Näin ollen jäsenvaltioiden on tarkistettava syrjinnän vastaista lainsäädäntöään ja muutettava sitä tarvittaessa, jotta voidaan varmistaa, että se kattaa kaikki tapaukset, joissa tekoäly- ja ADM-järjestelmien käyttö voi aiheuttaa välitöntä tai välillistä syrjintää, mukaan lukien läheissyrjintää ja intersektionaalista syrjintää.

2.4.3. Tietojen keräämistä ja käsittelyä koskevat rajoitukset

Yhdenvertaisuustietojen keräämistä koskevat nykyiset oikeudelliset rajoitukset rajoittavat mahdollisuuksia tutkia tehokkaasti algoritmista syrjintää ja edistää yhdenvertaisuutta tekoäly- ja ADM-järjestelmien käytössä. Jotta tekoälyyn liittyvään syrjintään voitaisiin puuttua tehokkaammin, saatetaan tarvita siihen paremmin soveltuvia yhdenvertaisuustietojen keräysmenettelyjä. Muut täytäntöönpanokysymykset liittyvät arkaluonteisten henkilötietoryhmien käyttöön. Tekoälyasetuksen 10 artiklan 5 kohdassa sallitaan poikkeuksellisesti arkaluonteisten henkilötietoryhmien, kuten etnisen alkuperän tai seksuaalisen suuntautumisen, käsittely (yleisen tietosuoja-asetuksen 9 artiklan määritelmän mukaisesti) vinoumien tunnistamista varten. Tämä mahdollisuus ei kuitenkaan koske tekoälyjärjestelmiä, joita ei luokitella suuririskisiksi. Tämä voi aiheuttaa merkittäviä rajoituksia syrjivien vinoumien havaitsemiselle, ehkäisemiselle ja lieventämiselle järjestelmissä, joita pidetään matalan riskin järjestelminä ja joihin yhdenvertaisuusvelvoitteita sovelletaan monialaisesti. On myös epävarmaa, voivatko yhdenvertaisuuselimet ja muut perusoikeuksia suojaavat viranomaiset hyödyntää yleisen tietosuoja-asetuksen 9 artiklassa ja tekoälyasetuksen 10 artiklan 5 kohdassa kuvattuja poikkeuksia tiettyjen tekoälyjärjestelmien syrjivien vaikutusten arvioinnissa. Algoritmista syrjintää voi olla vaikeampi havaita ja valvoa, kun arkaluonteisia henkilötietoryhmiä – jotka ovat päällekkäisiä monien (mutta eivät kaikkien) EU:n ja kansallisen yhdenvertaisuuslainsäädännön nojalla kiellettyjen syrjintäperusteiden kanssa – ei voida käyttää testaus-, tarkastus- ja arviointitarkoituksiin.

2.5. Tekoälyjärjestelmien ”riskien vähentäminen”

Tekoälyasetuksen riskiperusteista luokittelua on arvosteltu. Asetuksen liitteessä III olevaa luetteloa suuririskisistä sovelluksista on pidetty liian joustamattomana nopeasti kehittyvän teknologian kannalta. Tekoälyasetuksessa ei myöskään anneta selkeitä ohjeita sen määrittämiseksi, onko liitteessä III lueteltuihin aloihin kuuluva sovellus suuririskinen vai ei. Tämä on erityisen ongelmallista, kun otetaan huomioon useita poikkeuksia sisältävä tekoälyasetuksen 6 artiklan 3 kohta. Tarjoajat voivat esimerkiksi

ilmoittaa, että järjestelmät, jotka tavallisesti kuuluisivat suuririskisten järjestelmien luetteloon, eivät aiheutakaan merkittävää riskiä luonnollisten henkilöiden terveydelle, turvallisuudelle tai perusoikeuksille. Tekoälyjärjestelmiä ei pidetä suuririskisinä, kun niiden tarkoituksena on "suorittaa suppea menettelyllinen tehtävä" tai "valmistelutehtävä",²⁶ "parantaa aiemmin suoritettua ihmisen toiminnan tulosta" tai "havaita päätöksentekotapoja tai poikkeamia aiemmista päätöksentekotavoista, eikä sen tarkoituksena ole korvata aiemmin tehtyä ihmisen tekemää arviota tai vaikuttaa siihen ilman asianmukaista ihmisen suorittamaa arviota". Jos järjestelmä ei tarjoajan mielestä ole suuririskinen, tarjoajan on ainoastaan "dokumentoitava arvionsa ennen kuin kyseinen järjestelmä saatetaan markkinoille tai otetaan käyttöön" (6 artiklan 4 kohta) ja rekisteröitävä järjestelmä 49 artiklan 2 kohdan mukaisesti.

Tässä vaiheessa ei ole systemaattista valvontaa, sillä kansalliset toimivaltaiset viranomaiset voivat pyytää arviointeja, mutta niillä ei ole systemaattista velvollisuutta tehdä näin. Tämä luo mahdollisen porsaanreiän, jota saatetaan käyttää laajalti tekoälyjärjestelmien "riskin vähentämiseen", jotta välttyttäisiin noudattamasta tekoälyasetuksen mukaisia suuririskisiin järjestelmiin sovellettavia vaatimuksia ja suoja-toimia.²⁷ Tässä yhteydessä tekoälyä koskevan puiteyleissopimuksen lähestymistapa, joka perustuu oikeuksiin eikä riskeihin, voisi osoittautua erityisen hyödylliseksi perusoikeuksia koskevien suojatoimien laajentamisessa koskemaan tekoäly- ja ADM-sovelluksia, joiden "riskejä on vähennetty". Toinen asetusten mahdollinen aukko, jota saatetaan hyväksikäyttää, liittyy tekoälyn määritelmään. Koska molempia tekstejä sovelletaan tekoälyyn eikä esimerkiksi yksinkertaisiin sääntöihin perustuviin järjestelmiin, tietyt ADM-järjestelmät saatetaan jättää soveltamisalan ulkopuolelle, jos niiden ei katsota sisältyvän tekoälyn määritelmään (Euroopan komissio, 2025d). Tämä on toinen porsaanreikä, jota yksinkertaisempien ADM-järjestelmien tarjoajat voivat käyttää tekoälyasetuksen velvoitteiden kiertämiseen.

2.6. Tekniset standardit, yhdenvertaisuuselimet ja yhdenmukaistamiseen liittyvä kysymys

CEN-CENELEC julkaisee tekoälyä koskevat yhdenmukaistetut standardit, ja ne tulevat olemaan kriittisen tärkeä väline perusoikeuksien noudattamisen tukemisessa, mukaan lukien algoritmisen syrjinnän ehkäisemisessä ja siltä suojautumisessa. Keskeinen kysymys on kuitenkin se, mitä syrjintäriskejä standardit edellyttävät tarjoajien tunnistavan, arvioivan ja käsittelevän? Mitkä syrjinnän muodot tarjoajien on erityisesti otettava huomioon ja miten niitä on käsiteltävä, jotta niiden voidaan katsoa noudattavan tekoälyasetusta? Tähän liittyy useita ongelmia. Ensimmäinen niistä koskee yhteissääntelyyn osallistuville yksityisille organisaatioille annettua valtaa. Useimmat CEN-CENELECin JTC 21 -komiteassa mukana olevat organisaatiot, jotka vastaavat näiden standardien laatimisesta, ovat alan toimijoita, joilla on resursseja rahoittaa tällainen osallistuminen. Kansalaisjärjestöt ja organisaatiot, joiden oikeutettu etu on kyseessä, voivat osallistua jossain määrin (esimerkiksi Equinet osallistuu tällä hetkellä standardien laatimisprosessiin), mutta niillä on usein vain rajalliset taloudelliset

26. Kuitenkin 6 artiklan 3 kohdan d alakohdan mukaan tekoälyjärjestelmän käyttö luonnollisten henkilöiden profilointiin liitteessä III luetelluilla alueilla on aina suuririskistä.

27. Nämä vaatimukset ja suojatoimet on lueteltu suuririskisiä tekoälyjärjestelmiä koskevassa osiossa.

resurssit siihen. Tämä tarkoittaa myös sitä, että käytännön asiantuntemus voi jakautua epätasaisesti: teknistä asiantuntemusta on riittävästi mutta oikeudellista ja varsinkin ihmisoikeuksia koskevaa asiantuntemusta vähemmän. Standardiluonnokset eivät ole julkisia, mikä vaikeuttaa parhaillaan käsiteltävänä olevien täytäntöönpanostrategioiden tehokkuuden ja puutteiden seuranta- ja arviointia. Tämä ehkäisee yrityksiä vaikuttaa näihin keskusteluihin ja käsitellä nykyisiä rajoituksia.

Toinen ongelma on tämän lähestymistavan yhteensopivuus ja täydentävyys EU:n perusoikeuslainsäädännön, mukaan lukien yhdenvertaisuuslainsäädännön, kanssa. Toisaalta on vielä epäselvää, kuinka kattavaa riskinhallintaa standardit edellyttävät vaatimustenmukaisuuden varmistamiseksi. Tämä on keskeinen kysymys, koska tekoälyasetuksen 9 artiklan 5 kohdassa todetaan, että ”riskinhallintatoimenpiteiden on oltava sellaisia, että kuhunkin vaaraan liittyvän asiaankuuluvan jäännösriskin ja suuririskisten tekoälyjärjestelmien kokonaisjäännösriskin katsotaan olevan hyväksyttävällä tasolla”. Vinoumien ja syrjinnän osalta tekoälyasetuksen mukaiset hyväksyttävät jäännösriskit eivät takaa EU:n perusoikeuslainsäädännön noudattamista, sillä siinä käytetään erilaista lähestymistapaa, jossa keskitytään todellisiin haittoihin. Itse asiassa EU:n perusoikeuskirjan 52 artiklan 1 kohdassa todetaan, että ”[t]ässä perusoikeuskirjassa tunnustettujen oikeuksien ja vapauksien käyttämistä voidaan rajoittaa ainoastaan lailla, ja kyseisten oikeuksien ja vapauksien olennaista sisältöä noudattaen” ja ”[s]uhteellisuusperiaatteen mukaisesti rajoituksia voidaan tehdä ainoastaan, jos ne ovat välttämättömiä ja vastaavat tosiasiallisesti unionin tunnustamia yleisen edun mukaisia tavoitteita tai tarvetta suojella muiden henkilöiden oikeuksia ja vapauksia”. Sallittujen häiriöiden – joita rajoittavat perusoikeuksien olemus ja joiden edellytyksenä ovat välttämättömyys ja yleinen etu – laajuus näyttää olevan ristiriidassa tekoälyasetuksessa käytetyn jäännösriskiä soveltavan lähestymistavan kanssa. Näiden käsitteiden välinen suhde on vähintäänkin epäselvä ja vaatii selvennystä, mahdollisesti unionin tuomioistuimelta.

Toisaalta EU:n käyttämä standardointilähestymistapa yhdistettynä EU:n uuteen tekoälyasetukseen sisältyvään (ainakin osittaiseen) enimmäistason yhdenmukaistamisen kehykseen on ristiriidassa EU:n syrjinnän vastaisten direktiivien vähimmäistason yhdenmukaistamisen lähestymistavan kanssa. Vaikka EU:n sekundaarilainsäädäntö kieltää kuusi syrjintäperustetta (rotu ja etninen alkuperä, biologinen ja sosiaalinen sukupuoli, vammaisuus, seksuaalinen suuntautuminen, uskonto tai vakaumus ja ikä), se antaa jäsenvaltioille myös mahdollisuuden mennä näitä vähimmäisvaatimuksia pidemmälle ja kieltää syrjintä laajemmin niin sanotuilla ”suotuisammilla määräyksillä”, kunhan ne ovat EU:n sopimusten mukaisia. On kuitenkin epäselvää, millaisia vinoumamuotoja suuririskisten järjestelmien tarjoajien on testattava. Jos esimerkiksi vinoumien testausta ja lieventämistä edellytetään vain kuuden EU:n tasolla kielletyn syrjintäperusteen osalta ja näitä standardeja noudattavien tarjoajien oletetaan sitten noudattavan myös tekoälyasetusta, tarjoajat voivat tällöin myydä tuotteitaan koko sisämarkkinoilla. Vaikka käyttöönottajat voivat edelleen olla vastuussa syrjinnästä EU:n ja kansallisten laajempaa suojaa edellyttävien syrjinnän vastaisten lakien nojalla, kysymyksiä herää ainakin kahdessa suhteessa.

Ensinnäkin EU:n syrjintälainsäädäntö sallii nimenomaisesti suotuisammat kansalliset syrjintää koskevat määräykset, eikä niitä tulisi pitää vapaan liikkuvuuden esteenä. Tällainen lähestymistapa rikkoo myös EU:n perusoikeuskirjan 21 artiklan 1 kohtaa,

jossa "[k]ielletään kaikenlainen syrjintä, joka perustuu sukupuoleen, rotuun, ihonväriin tai etniseen taikka yhteiskunnalliseen alkuperään, geneettisiin ominaisuuksiin, kieleen, uskontoon tai vakaumukseen, poliittisiin tai muihin mielipiteisiin, kansalliseen vähemmistöön kuulumiseen, varallisuuteen, syntyperään, vammaisuuteen, ikään tai sukupuoliseen suuntautumiseen tai muuhun sellaiseen seikkaan". Toinen vaihtoehto olisi käyttää perusoikeuskirjaan sisältyvää syrjinnän laajempaa määritelmää. Jännitteet laajempaa suojaa tarjoavan kansallisen lainsäädännön, joka on nimenomaisesti sallittu EU:n syrjintälainsäädännön nojalla, kanssa säilyisivät kuitenkin. Jos standardeissa ei kuitenkaan mainita kiellettyjä syrjintäperusteita, jotka suuririskisten tekoöljyjärjestelmien tarjoajien on otettava huomioon, kun ne testaavat vinoumia tai arvioivat järjestelmänsä vaikutusta perusoikeuksiin, on olemassa merkittävä riski, että tällaiset menettelyt jäävät suurelta osin tehottomiksi.

Toinen kysymys liittyy siihen, miten tällaiset vaatimustenmukaisuutta koskevat mekanismit vaikuttavat todistustaakan siirtymiseen syrjintätapauksissa. Tässä yhteydessä on tärkeää muistaa, että tekoöljyasetuksen 40 artiklan mukainen oletama vaatimusten noudattamisesta koskee vain asetuksen mukaisten suuririskisten järjestelmien vaatimuksia (katso riskinhallintajärjestelmää koskeva osio), ei syrjinnän vastaista tai muuta perusoikeuslainsäädäntöä. Tältä osin tekoöljyasetuksen 82 artiklan 1 kohdassa tunnustetaan kyllä, että "vaikka suuririskinen tekoöljyjärjestelmä on tämän asetuksen mukainen, se [voi] kuitenkin aiheuttaa riskin ihmisten terveydelle tai turvallisuudelle, perusoikeuksille tai muille yleisen edun suojaan liittyville näkökohdille".

Näin ollen on korostettava, että tekoöljyasetuksen (osittaista) enimmäistason yhdenmukaistamisen lähestymistapaa ei tulisi tulkita siten, että se kumoaisi tai vähentäisi EU:n primaari- ja sekundaarilainsäädännön olemassa olevia syrjinnän vastaisia velvoitteita.

Mitä tämä tarkoittaa yhdenvertaisuuselimille ja perusoikeuksia suojeleville viranomaisille, kuten tekoöljyasetuksen 77 artiklan mukaan nimetyille viranomaisille? Ensinnäkin voi olla vaikeaa arvioida riskinhallintajärjestelmien ja perusoikeusvaikutusten arviointien syrjinnän vastaisen lainsäädännön noudattamista, jos niissä ei anneta tietoja testatuista syrjintäperusteista. Pakottava kysymys on, onko yhdenvertaisuuselimillä ja perusoikeuksia suojaavilla sidosryhmillä riittävän tarkat tiedot sen arvioimiseksi, miten syrjintää on ehkäisty ja/tai riskejä lievennetty oikeudellisesti suojattujen ryhmien osalta suuririskisten tekoöljyjärjestelmien yhteydessä ja miten nämä ryhmät on määriteltävä riskinarviointimenettelyjen aikana. Toiseksi edellä mainittu oikeudellinen ristiriita luo kuilun tarjoajien (ja tapauksen mukaan myös käyttöönottajien) velvollisuuksien, jotka koskevat syrjiviin vinoumiin riskien ehkäisemistä ja torjumista suuririskisissä järjestelmissä, ja yhdenvertaisuuselinten ja muiden asiaankuuluvien viranomaisten toimivaltuuksien välille. Useimmilla yhdenvertaisuuselimillä on valtuudet käsitellä syrjintää, joka perustuu kaikkiin kansallisessa lainsäädännössä kiellettyihin syrjintäperusteisiin. Nämä perusteet usein ylittävät EU:n syrjinnän vastaisessa sekundaarilainsäädännössä kielletyt perusteet. Uusissa standardeissa tai niiden soveltamisessa suosittavasta lähestymistavasta (esimerkiksi pienin yhteinen nimittäjä tai määrittelemättömät vinoumatyypit) riippuen yhdenvertaisuuselinten on kiinnitettävä erityistä huomiota syrjintäperusteisiin, jotka on kielletty ainoastaan kansallisessa lainsäädännössä. Tarjoajat eivät välttämättä ole testanneet algoritmisen syrjinnän riskejä näiden kansallisesti kiellettyjen syrjintäperusteiden osalta.

Lisäksi standardointi luo yhdenvertaisuuselimille useita erilaisia haasteita. Yhdenvertaisuuselinten voi olla tärkeää seurata, miten CEN-CENELEC määrittelee uudet yhdenmukaistetut tekoälystandardit, ja vaikuttaa niiden laadintaan, koska näiden standardien noudattaminen johtaa siihen, että suuririskisten tekoälyjärjestelmien tarjoajien oletetaan täyttävän tekoälyasetuksen olennaiset vaatimukset. Tähän liittyy kuitenkin useita ongelmia. Ensinnäkin, vaikka Equinet on mukana tietyissä tekoälystandardien laatimisesta vastaavan JTC 21 -komitean alaryhmissä, sovitujen standardien julkisuuden ja saatavuuden aste on edelleen epäselvä. Asiassa *Public.Resource.Org ja Right to Know v. komissio ym.* unionin tuomioistuin päätti, että standardit on saatettava julkisesti EU:n kansalaisten saataville (Euroopan unionin tuomioistuin, 2024). Asia koski lelujen turvallisuuteen liittyvien standardien saatavuutta, joten on edelleen epäselvää, sovelletaanko tätä avoimuusvaatimusta myös tekoälystandardeihin ja asetetaanko ne maksutta saataville.

Toiseksi saatetaan tarvita teknistä asiantuntemusta, jotta voidaan ymmärtää ja arvioida kriittisesti, varmistavatko tekoälystandardit riittävästi perusoikeuksien ja syrjinnän vastaisen lainsäädännön noudattamisen. Equinetin raportin mukaan standardit "määrittelevät, mitä tietoja ja kuinka yksityiskohtaisia asiakirjoja dokumentaation sisällytetään" tekoälyasetuksen mukaisten suuririskisten tekoälyjärjestelmien dokumentaatiota koskevien olennaisten vaatimusten mukaisesti. Näin ollen standardit myös määrittävät, "millaista riskinhallintaa ja markkinoille saattamisen jälkeistä seuranta tarjoajien on tehtävä perusoikeuksiin kohdistuvien riskien ennakointia, tunnistamista ja lieventämistä varten" (Mittelstadt, 2025).

Yhdenvertaisuuselinten olisi siis kiinnitettävä erityistä huomiota uusien tekoälystandardien teknisiin valintoihin:

- ▶ Mitä oikeudenmukaisuusmittareita käytetään suorituskyvyn mittaamiseen? Millaisia kynnysarvoja käytetään syrjivien vinoumien riskien kvantifointiin? Mitä vinoumia poistavia toimenpiteitä ja oikeudenmukaisuusmenetelmiä käytetään? Ovatko käytetyt menetelmät yhdenmukaisia EU:n syrjinnän vastaisen lainsäädännön ja oikeuskäytännön kanssa?
- ▶ Miten syrjinnän "jäännösriskien" hyväksyttävät tasot määritellään?
- ▶ Miten järjestelmien käyttötarkoitus, odotettu tarkkuustaso ja ennakoitavissa olevat väärinkäytökset määritetään?
- ▶ Miten syrjintäriskejä ja -vaikutuksia mitataan ja lievennetään EU:n syrjinnän vastaisen lainsäädännön nojalla suojattujen ryhmien osalta?
- ▶ Miten tietyt ryhmät määritellään esimerkiksi teknisessä dokumentaatiossa, riskinhallinnassa ja vaikutusarvioinneissa?
- ▶ Onko intersektionaalinen syrjintä otettu riittävästi huomioon?
- ▶ Kattavatko ryhmien tekniset määritelmät koko kielletyillä syrjintäperusteilla suojatun väestöosan?

Näiden kysymysten huomioiminen on tärkeää etiikka- ja perusoikeuspesun estämisessä.

Tämä luo haasteen yhdenvertaisuuselimille, ja niiden voi olla tarpeen etsiä uutta asiantuntemusta, osaamista ja resursseja voidakseen vastata näihin kysymyksiin. Vankkojen yhteistyömallien luominen asiaankuuluvien viranomaisten

(markkinavalvontaviranomaiset, tekoälytoimisto ja niin edelleen), täydentävää asiantuntemusta omaavien elinten (tietosuojaviranomaiset, oikeusasiamiehet, kuluttajansuojavirastot ja niin edelleen) ja asiaankuuluvien kolmansien osapuolten kanssa (tutkijat, kansalaisjärjestöt ja niin edelleen) voi auttaa yhdenvertaisuuselimää ratkaisemaan asiantuntemukseen ja resursseihin liittyviä haasteita.

Tiivistelmä – tekoälyn käyttö Belgian julkishallinnossa ja algoritmista syrjintää koskevat poliittiset suositukset



Tiivistelmä perustuu nykytilaa ja puutteita analysoivaan raporttiin, jonka Belgian osuuden on laatinut Ine van Zeeland.²⁸

Tekoälyn käyttö julkishallinnossa

Belgian julkishallinnossa käytetään useita tekoälysovelluksia, ja käynnissä on useita tekoälyhankkeita esimerkiksi poliisi-, rahoitus-, työllisyys- ja koulutusaloilla. Koska näistä järjestelmistä ei ole kattavaa julkista rekisteriä, tekoälyn ja automaattisen päätöksenteon (ADM) järjestelmien käyttöä ei ole mahdollista kartoittaa täysin ja johdonmukaisesti Belgian julkisissa palveluissa. Belgian liittovaltion palveluiden hallintoelin FPS BoSA (Federal Public Service Policy and Support) isännöi tekoälyn seurantakeskusta, joka on käynnistänyt pyrkimykset kerätä tietoa tekoälyn tai ADM-järjestelmien käytöstä Belgian julkisissa palveluissa. Myös muut elimet, kuten Brysselin FARI-instituutti ja Belgian sosiaaliturvasta vastaava FPS Social Security, ovat toteuttaneet tai suunnittelevat vastaavia toimia tietojen keräämiseksi.

28 Tämä on luottamuksellinen raportti, jossa tarkastellaan tekoälyn ja automaattisten päätöksentekojärjestelmien käyttöä Belgian julkisella sektorilla, arvioidaan tekoälyyn liittyviä oikeudellisia ja poliittisia puitteita tasa-arvon ja yhdenvertaisuuden näkökulmasta sekä annetaan suosituksia viranomaisille ja päätöksentekijöille.

Lainsäädäntö ja asiaankuuluva oikeuskäytäntö

EU:n tekoälyasetus yhdessä eurooppalaisen ja belgialaisen syrjinnän vastaisen ja tietosuojaalainsäädännön kanssa muodostaa oikeudellisen kehyksen algoritmiseen syrjintään puuttumiselle. Tekoälyyn, ADM-järjestelmiin ja syrjintään liittyvä oikeuskäytäntö on Belgiassa suhteellisen tuoreta. Nykytilaa ja puutteita analysoivassa raportissa tunnistettiin muutamia aiheeseen liittyviä tapauksia, kuten unionin tuomioistuimen päätös asiassa C-817/19 *Ligue des droits humains ASBL v. Conseil des ministres*,²⁹ jossa keskityttiin koneoppimisalgoritmien avulla tapahtuvan automaattisen tietojenkäsittelyn ja syrjimättömyyden väliseen jännitteeseen.

Tietosuojaan liittyen Belgian tietosuojaviranomaisen GBA-APD:n riita-asioita käsittelevä jaosto on antanut kaksi tuomiota ADM-järjestelmiä koskevissa tapauksissa, joilla voi olla merkitystä syrjinnän torjunnan kannalta. Ensimmäinen tapaus koski autojen yhteiskäyttöyrityksen luottopisteytystä³⁰ ja toinen maksukorttitietoihin perustuvan algoritmisen mallin kehittämistä.³¹ Molemmissa tapauksissa riita-asioita käsittelevän jaoston päätökset koskivat kuitenkin yleisen tietosuoja-asetuksen vaatimuksia, jotka eivät liittyneet suoraan syrjintään. On mahdollista, että kaikkia tuomioistuinten käsittelyyn sisältyviä, tekoäly- ja ADM-järjestelmiin liittyvää eriarvoista kohtelua koskevia tapauksia ei ole havaittu. Vuonna 2022 Brysselin ensimmäisen oikeusasteen tuomioistuin antoi esimerkiksi tuomion asiassa, joka koski välillistä syrjintää pysäköintipalvelun rekisterikilpiä lukevan ADM-järjestelmän kautta.³²

EU:n säännösten lisäksi mahdollisuuksia algoritmisen syrjinnän ehkäisemiseen ja siihen puuttumiseen tarjoaa syyskuussa 2024 allekirjoitettavaksi tullut tekoäly ja ihmisoikeuksia, demokratiaa ja oikeusvaltioperiaatetta koskeva Euroopan neuvoston puiteyleissopimus, jonka tasa-arvoa ja syrjimättömyyttä koskevassa 10 artiklassa kielletään syrjintä ja edellytetään, että kukin valtio "sitoutuu ottamaan käyttöön tai pitämään voimassa toimenpiteet, joilla pyritään poistamaan eriarvoisuutta ja saavuttamaan oikeudenmukaisia ja tasapuolisia tuloksia".

Oikeudellisten ja poliittisten kehysten puutteet

Avoimuuden puute

Kriittinen ongelma Belgiassa on tekoäly- ja ADM-järjestelmien käytön avoimuuden ja raportoinnin puute, sillä tällä hetkellä julkisen sektorin organisaatioilla ei ole virallista velvoitetta rekisteröidä tekoäly- ja ADM-järjestelmien käyttöä tai ilmoittaa siitä. Kattavan ja julkisesti saatavilla olevan tietokannan puuttuminen aiheuttaa riskejä sekä valvonnalle että vastuuvollisuudelle, minkä lisäksi yhdenvertaisuuselimille syntyy haasteita seurata syrjintäriskejä tai tukea henkilöitä, jotka voivat kokea tällaisten

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

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. Asia 21/6495/C, saatavilla osoitteessa www.unia.be/fr/legislation-et-jurisprudence/jurisprudence/tribunal-de-premiere-instance-de-bruxelles-le-2-mai-20221, viitattu 10. marraskuuta 2025. Tapauksen kantajina olivat toimintarajoitteinen henkilö ja Collectif accessibilité Wallonie Bruxelles -järjestö (CAWaB) yhdenvertaisuustyöhön erikoistuneen Unia-instituution tuella.

järjestelmien aiheuttamaa syrjintää. Alan tutkimuksissa on jatkuvasti havaittu, että algoritmista syrjintää tapahtuu raportoitua enemmän, eikä syrjintää tutkita riittävästi.

Perusoikeuksia koskevan tietoisuuden ja yhdenmukaisen lähestymistavan puute

Erilaisista koulutuksista ja eettisistä ohjeista huolimatta monissa julkisissa palveluissa on vain vähän systemaattista tietoa syrjintään liittyvistä riskeistä. Julkishallinnolta puuttuu yleisesti ottaen standardoitu lähestymistapa perusoikeuksiin ja syrjintään liittyvien riskien arviointiin tekoälyn käytössä julkisissa palveluissa. Lisäksi puuttuu yhdenmukainen lähestymistapa, jolla tekoäly- ja ADM-järjestelmien suunnittelussa, kehittämisessä ja käyttöönotossa voitaisiin ottaa huomioon asianomaisten ryhmien tai yksilöiden ja perusoikeusasiantuntijoiden näkökulmat.

Hajanainen valvontaympäristö

Belgiassa on institutionaalisesti monimutkainen valvontaympäristö, jossa valtuudet jaetaan liittovaltion, alueiden ja yhteisöjen tasojen välillä. Tämä luo haasteen valvonnan ja oikeudellisten tulkintojen yhdenmukaistamiselle ja aiheuttaa riskin siitä, että tekoälyasetusta sovelletaan epä johdonmukaisesti Belgian julkisissa palveluissa.

Resurssien ja asiantuntemuksen puutteet

Yhdenvertaisuuselimillä ja valvontaviranomaisilla on Belgiassa huomattavia haasteita käsitellä tekoälyyn ja ADM-järjestelmiin perustuvaa syrjintää, koska julkiset taloudelliset, tekniset ja henkilöstöresurssit ovat rajalliset.

Keskeiset suositukset

- ▶ **Tekoäly- ja ADM-järjestelmien julkisen rekisterin perustaminen:** Euroopan laajuisen rekisterin lisäksi voidaan ottaa käyttöön lakisääteinen velvoite ilmoittaa tekoäly- ja ADM-järjestelmien käytöstä julkisella sektorilla kansalliselle ja ainakin osittain julkisesti saatavilla olevalle rekisterille tai tietopankille, jotta voidaan varmistaa avoimuus ja tehokas valvonta.
- ▶ **Eri tasojen toimivaltaisten kansallisten viranomaisten edistyneiden koordinoitimekanismien käyttö:** jotta voidaan varmistaa, että tekoälyasetuksen 77 artiklan mukaiset viranomaiset ja kansalliset ihmisoikeusrakenteet otetaan ennakoivasti mukaan tutkimuksiin, on suositeltavaa tehdä ja kannustaa viranomaisia tekemään virallisia yhteistyösopimuksia, joilla varmistetaan asianomaisten viranomaisten yhdenmukainen valvonta ja päätöksenteko, sekä luoda suuntaviivoja kattavaa osallistumista varten.
- ▶ **Verkoston perustaminen lupaavien käytäntöjen jakamista ja synergian lisäämistä varten:** Viranomaisia on suositeltavaa tukea liittovaltion rakenteellisen yhteistyöverkoston perustamisessa, jotta käytännön neuvoja ja ohjeita sekä hyviä käytäntöjä voidaan jakaa eri viranomaisten välillä. Näin voidaan puuttua tekoäly- ja ADM-järjestelmien, jotka jo lähtökohtaisesti suojaavat perusoikeuksia tai ainakin mahdollistavat vaikutusten seurannan koko järjestelmän elinkaaren ajan, kehittämiseen tai hankintaan tarvittavan asiaankuuluvan tietämyksen

ja valmiuksien pirstaleisuuteen. Tällaiseen verkostoon on suositeltavaa sisällyttää syrjinnän torjunnan, perusoikeuksien ja henkilötietojen suojaamisen asiantuntijoita.

- ▶ **Menettelyllisten takeiden varmistaminen, jotta oikeussuojakeinot olisivat tehokkaampia tekoölyyn tai ADM-järjestelmiin liittyvissä tapauksissa, mukaan lukien syrjinnän vastaisen lainsäädännön päivittäminen tarvittaessa:** Todistustaakkaa voidaan siirtää lainsäädännöllä ottamalla käyttöön algoritmisen vinouman olettama. Tämä voi auttaa henkilöitä, joille todisteiden esittämiseen liittyy muutoin lähestulkoon mahdottomia esteitä. Tarvittaessa syrjinnän vastaista lainsäädäntöä on muutettava tämän olettamuksen tukemiseksi. Samalla on helpotettava yleiseen etuun liittyvien järjestöjen mahdollisuuksia käynnistää kollektiivisia oikeussuojajamenettelyjä esimerkiksi työllisyyden, terveydenhuollon ja sosiaaliturvan aloilla, kuten kuluttaja-asioihin liittyvissä sääntely- ja valitusmenettelyissä on jo yleistä.
- ▶ **Perusoikeusvaikutusten arviointien yhteenvetojen laatiminen ja julkaisu:** Perusoikeusvaikutusten arviointeihin (FRIA) on suositeltavaa soveltaa yhdenmukaista lähestymistapaa. Näitä arviointeja on käytettävä kaikissa julkisen sektorin organisaatioissa, ja niissä on otettava huomioon asianomaisten ryhmien ja yhdenvertaisuuselinten näkökulmat. FRIA-arviointien yhteenvedot on myös julkaistava.
- ▶ **Syrjimättömyyden sisällyttäminen tekoöly- ja ADM-järjestelmien julkisiin hankintoihin:** on varmistettava, että tekoölyn ja ADM-järjestelmien julkisiin hankintoihin sisällytetään yhdenvertaisuuden vaatimukset asianmukaisen huolellisuusvelvoitteen noudattamisen parantamiseksi.
- ▶ **Tekoölyn käyttöä koskevan tietoisuuden ja tekoölylukutaidon lisääminen:** yleisön tietoisuutta tekoölyn käytöstä julkisissa palveluissa voidaan lisätä tiedotuskampanjoilla ja tekoölylukutaitoa edistävillä toimilla esimerkiksi koulutuksessa.
- ▶ **Tekoölyn käyttöön liittyvien riittävien julkisten resurssien ja asiantunteumuksen saatavuuden takaaminen yhdenvertaisuuselimille ja valvontaviranomaisille:** nämä ovat välttämättömiä yhdenvertaisuuselimille ja valvontaviranomaisille, jotta ne voivat täyttää tekoölyasetuksen ja syrjinnän vastaisen lainsäädännön mukaiset tehtävänsä.
- ▶ **Julkisen sektorin henkilöstön koulutus tekoöly- ja ADM-järjestelmien vaikutuksista perusoikeuksiin:** koulutuksessa on käsiteltävä tekoölyn perusoikeusvaikutuksia eettisiä näkökohtia laajemmin ja kiinnitettävä erityistä huomiota syrjintäriskeihin.
- ▶ **Yhdenvertaisuuselinten, muiden julkisten instituutioiden ja kansalaisyhteiskunnan välisen yhteistyön vahvistaminen:** yhteistyötä voidaan vahvistaa ja vakiinnuttaa esimerkiksi ”käytäntöyhteisöillä”, jotka vähentävät asianomaisten henkilöiden esteitä hakea tukea tarvittaessa.

Tiivistelmä – tekoälyn käyttö Suomen julkishallinnossa ja algoritmista syrjintää koskevat poliittiset suositukset



Tiivistelmä perustuu nykytilaa ja puutteita analysoivaan raporttiin, jonka Suomen osuuden on laatinut Emeline Banzuzi.³³

Automaattisen päätöksenteon vakiintunut käyttö ja tekoälyn kehittyvä käyttö julkishallinnossa

Suomessa julkishallinnossa on jo pitkään käytetty automaattisen päätöksenteon (ADM) järjestelmiä, joiden avulla voidaan tehdä tehokkaasti erilaisia hallinnollisia päätöksiä erityisesti muuttoliikettä, poliisitoimintaa, työllisyyttä ja rekrytointia koskevissa asioissa. Sen sijaan tekoälyn (AI) käyttö on edelleen vähäistä ja rajoittuu enimmäkseen generatiivisen tekoälyn (GenAI) käyttöön tehokkuuden parantamiseksi, kuten Microsoft Copilotin käyttöön asiakirjojen laatimisen tai muistiinpanojen tekemisen tukena. Useat viranomaiset kuitenkin tutkivat tai kokeilevat tekoälysovellusten käyttöä. Maahanmuuttovirasto (Migri) esimerkiksi harkitsee tekoälyn käyttöä puheesta tekstiksi -kääntämisessä, Suomen poliisi käyttää tekoälyä rekisterikilpien lukemiseen ja Helsingin kaupunki kokeilee "Helsinki GPT" -tekoälytyökalua asiakaspalautteiden automaattiseen luokitteluun ja priorisointiin sekä vastausten tuottamiseen.

33 Tämä on luottamuksellinen raportti, jossa tarkastellaan tekoälyn ja automaattisten päätöksentekojärjestelmien käyttöä Suomen julkisella sektorilla, arvioidaan tekoälyn liittyviä oikeudellisia ja poliittisia puitteita tasa-arvon ja yhdenvertaisuuden näkökulmasta sekä annetaan suosituksia viranomaisille ja päätöksentekijöille.

Oikeudellinen ja institutionaalinen kehys

EU:n tekoälyasetus yhdessä eurooppalaisen ja suomalaisen syrjinnän vastaisen ja tietosuojaalainsäädännön kanssa muodostaa oikeudellisen kehyksen algoritmiseen syrjintään puuttumiselle. Suomen viranomaiset valmistautuvat mukauttamaan oikeudellisia ja institutionaalisia kehyksiään tekoälyasetuksen mukaisiin velvoitteisiin. Tekoälyasetuksen täytäntöönpanoa varten Suomessa on vireillä kaksi lainsäädäntöehdotusta. Ensimmäisellä pyritään esimerkiksi sääntelemään tiettyjen tekoälyjärjestelmien valvontaa ja rikkeistä määrättäviä sakkoja sekä nimeämään kansallisessa lainsäädännössä toimivaltaiset viranomaiset ja yksi yhteyspiste tekoälyasetuksen edellyttämällä tavalla. Toisella pyritään sääntelemään esimerkiksi ”tekoälyn testiympäristöjen” perustamista ja toimintaa. Muihin institutionaalsiin valmistelutöihin kuuluu koordinoitumekanismien luominen ja valvontatehtävien selventäminen.

Suomessa ADM-järjestelmien käytöstä säädetään erikseen hallintolaissa. Sen mukaan ADM-järjestelmän käyttö on sallittua vain tilanteissa, joissa päätöksen tekeminen ei edellytä tapauskohtaista harkintaa tai joissa virkamies ratkaisee tapauskohtaista harkintaa edellyttävän seikan. ADM-järjestelmän on perustuttava luonnollisen henkilön ennalta määrittämiin sääntöihin.

Suomen yhdenvertaisuus- ja yhdenvertaisuuslainsäädäntö on suhteellisen kattava, sillä lainsäädännöllä vahvistetaan yhdenvertaisuuden edistämistä sen sijaan, että keskityttäisiin pelkästään syrjinnän ehkäisemiseen. Yhdenvertaisuuden edistämismahdollisuus edellyttää, että viranomaiset varmistavat yhdenvertaisen kohtelun aktiivisilla ja ennakoivilla toimenpiteillä, ja tämä velvoite koskee myös tekoäly- ja ADM-järjestelmien suunnittelua, kehittämistä ja käyttöä. EU:n säännösten lisäksi mahdollisuuksia algoritmisen syrjinnän ehkäisemiseen ja siihen puuttumiseen tarjoaa syyskuussa 2024 allekirjoitettavaksi tullut ja EU:n allekirjoittama tekoäly ja ihmisoikeuksia, demokratiaa ja oikeusvaltioperiaatetta koskeva Euroopan neuvoston puiteyleissopimus, jonka tasa-arvoa ja syrjimättömyyttä koskevassa 10 artiklassa kielletään syrjintä ja edellytetään, että kukin sopimuspuoli ”sitoutuu ottamaan käyttöönsä tai pitämään voimassa toimenpiteet, joilla pyritään poistamaan eriarvoisuutta ja saavuttamaan oikeudenmukaisia ja tasapuolisia tuloksia”.

Ainoa laajalti tunnettu syrjintää ja ADM-järjestelmiä koskeva tuomio Suomessa on rahoituspalveluyritystä koskeva tapaus. Siinä Suomen yhdenvertaisuus- ja tasa-arvolautakunta määräsi yrityksen lopettamaan tilastollisen luottopisteytysmenetelmän käytön, sillä se perustui nimenomaisesti asuinpaikkaan, ikään ja sukupuoleen eli kiellettyihin syrjintäperusteisiin.³⁴

34. Suomen yhdenvertaisuus- ja tasa-arvolautakunnan tapausseleste 216/2017. Saatavilla osoitteessa www.yvtltk.fi/material/sites/ytaltk/tapausselosteet/zedi26zya/Decision_YVTltk_xx16_2017_Multiple_discrimination_in_the_assessment_of_creditworthiness.pdf, viitattu 10. marraskuuta 2025.

Oikeudellisten ja poliittisten kehysten puutteet

Yhdenvertaisuus- ja syrjimättömyysvaikutusten arvioinnin standardoidun lähestymistavan tarve

Raportissa on havaittu, että tekoäly- tai ADM-järjestelmiä koskevia yhdenvertaisuus- ja syrjimättömyysvaikutusten arviointeja ei tehdä järjestelmällisesti, vaikka viranomaisia koskee lakisääteinen velvoite edistää yhdenvertaisuutta ja arvioida yhdenvertaisuusvaikutuksia. Suomen hallituksen selvitys- ja tutkimustoiminnan puitteissa kehitetty kansallinen arviointikehys³⁵ tarjoaa kuitenkin olemassa olevan välineen tekoälyjärjestelmien arviointiin. Yhdenvertaisuus- ja perusoikeusvaikutusten arviointia (FRIA) varten tarvitaan yhteinen menetelmä, jotta voidaan varmistaa, että yhdenvertaisuusnäkökohdat otetaan johdonmukaisesti huomioon tekoäly- ja ADM-järjestelmien koko elinkaaren ajan. Yhdenvertaisuusvaltuutettu on kehottanut hallitusta varmistamaan, että viranomaisilla, julkiset hankinnat mukaan lukien, on tehokas ja systemaattisesti sovellettava menetelmä yhdenvertaisuusvaikutusten arviointia varten.³⁶

Syrjintäriskien kapea-alainen tiedostaminen

Vaikka viranomaiset ovat saaneet koulutusta ja laatineet eettisiä ohjeita, lisäresurssit, kohdennettu lisäkoulutus ja syvällisempi ymmärrys algoritmisesta syrjinnästä ja syrjinnän vastaisesta lainsäädännöstä osana perusoikeusvelvoitteita olisivat silti hyödyllisiä.

Rajalliset resurssit sekä yhdenvertaisuuselinten ja valvontaviranomaisten ihmisoikeuksiin perustuvan teknisen asiantuntemuksen puute

Yhdenvertaisuuselinten ja muiden perusoikeuksien suojaamisesta vastaavien viranomaisten odotetaan saavan tekoälyasetuksen 77 artiklan nojalla entistä vahvempi rooli ja valtuudet tutustua dokumentaatioon ja pyytää suuririskisten järjestelmien testausta. Niiden kykyä täyttää tämä laajentunut tehtävä saattavat kuitenkin rajoittaa rajalliset tekniset, taloudelliset ja henkilöstöresurssit, sillä yleisön tietoisuus ja tekoälylukutaito ovat edelleen rajallisia, vaikka tekoälyjärjestelmien käyttö lisääntyykin. Lisäksi kaikilta toimijoilta, joilla on tekoälyasetuksen mukaisia valvontatehtäviä, edellytetään merkittävää tietämystä perusoikeuksien suojelusta, mukaan lukien syrjinnän eri muotojen tunnistamisesta ja arvioinnista. Resursseja olisi myös lisättävä, jotta voidaan varmistaa tehokas ennaltaehkäisevä mekanismi yhdenvertaisuuden turvaamiseksi ennen syrjintäriskien toteutumista.

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

36. "Non-Discrimination Ombudsman's recommendations to the governmental programme 2023–2027" (suomenkielinen alkuperäisteksti Yhdenvertaisuusvaltuutetun suositukset hallitusohjelmaan 2023–2027), saatavilla osoitteessa [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), viitattu 10. marraskuuta 2025.

Tekoäly- ja ADM-järjestelmiä koskevan julkisen ja yhdenmukaistetun rekisterin puute

Ei ole olemassa yhtä julkista kansallista tietokantaa, johon koottaisiin tietoja tekoäly- tai ADM-järjestelmien käytöstä Suomen julkishallinnossa. ADM-järjestelmien käyttöä koskevat kuitenkin hallintolain avoimuusveloitteet: viranomaisten on julkaistava ADM-järjestelmien käyttöä koskevia yleisiä tietoja verkkosivuillaan ja tiedotettava henkilöille, kun heitä koskevia päätöksiä on tehty ADM-järjestelmän avulla. Vaikka virastojen on julkaistava tietoja tällaisesta käytöstä omilla verkkosivustoillaan, nämä tiedot ovat edelleen hajanaisia.

Pirstaloitunut valvonta ja institutionaalinen koordinointi

Valvontavastuu on tällä hetkellä jaettu useille kansallisina toimivaltaisina viranomaisina toimiville viranomaisille, kuten markkina- ja perusoikeusviranomaisille. Vaikka tämä jako takaa kattavuuden eri sektoreilla, se aiheuttaa myös riskin epäjohtonmukaisista tulkinnoista ja täytäntöönpanoaukoista perusoikeuksien suojelun varmistamisessa, erityisesti jos yhteistyömekanismeja ei ole virallistettu ja viranomaisilla ei ole riittävästi yhdenvertaisuusasiantuntemusta.

Keskeiset suositukset

- **Syrjinnän vastaisen lainsäädännön, mukaan lukien yhdenvertaisuuden edistämismääräyksen, riittävän tietoisuuden ja täytäntöönpanon varmistaminen:** Tämä voidaan saavuttaa varmistamalla riittävä yhdenvertaisuusasiantuntemus, tarjoamalla koulutusta ja parantamalla toimivaltaisten viranomaisten tietoisuutta ja ymmärrystä näistä velvollisuuksista. Riittävää yhdenvertaisuusasiantuntemusta ja -opastusta on tarjottava myös henkilöille, jotka osallistuvat tekoälyn ostoon ja käyttöön liittyviin hankintoihin ja avustuksiin tai valvovat niitä.
- **Standardoitujen yhdenvertaisuus- ja perusoikeusvaikutusten arviointikehysten käyttö tekoälyjärjestelmien elinkaaren aikana:** Tällaisia arviointoja on suositeltavaa soveltaa kaikkien viranomaisten toiminnassa, ja yhdenvertaisuuselinten, kansalaisyhteiskunnan ja asianomaisten ryhmien on osallistuttava niihin. Arviointien tiivistelmät on myös julkaistava vastuuvetoisuuden lisäämiseksi.
- **Yhdenvertaisuus- ja valvontaviranomaisten resurssien lisääminen:** tekoälyasetuksen täytäntöönpanosta vastaaville kansallisille toimivaltaisille viranomaisille on suositeltavaa antaa riittävät resurssit, jotta ne voivat tehokkaasti hoitaa valvontatehtäviään ja osallistua merkityksellisesti tekoälyä koskevaan kansalliseen hallintokehykseen.
- **Tekoälyasetuksen mukaista EU:n laajuista tietokantaa täydentävän viranomaisten käyttämien tekoäly- ja ADM-järjestelmien kansallisen julkisen rekisterin perustaminen:** tällaisen rekisterin perustamisella voidaan lisätä avoimuutta, tukea valvontaa ja auttaa yhdenvertaisuuselimiä ja yleisöä seuraamaan syrjintäriskejä.

- **Toimivaltaisten viranomaisten välisten koordinointi- ja yhteistyömekanismien standardointi, myös yhdenvertaisuuden suojelun ja edistämisen osalta:** Toimivaltaisille viranomaisille on taattava riittävä yhdenvertaisuus- ja yhdenvertaisuusasiantuntemus, jotta voidaan varmistaa tekoälyasetuksen johdonmukainen täytäntöönpano. On tärkeää vahvistaa yhdenvertaisuusvaltuutetun ja tietosuojavaltuutetun välistä yhteistyötä algoritmisen syrjinnän saralla, koska tämän alan tapaukset voivat olla päällekkäisiä yhdenvertaisuuden ja rekisteröityjen tietosuojaoikeuksien kanssa.

Tiivistelmä – tekoälyn käyttö Portugalin julkishallinnossa ja algoritmista syrjintää koskevat poliittiset suositukset



Tiivistelmä perustuu nykytilaa ja puutteita analysoivaan raporttiin, jonka Portugalin osuuden on laatinut Eduardo dos Santos.³⁷

Tekoälyn käyttö julkishallinnossa

Portugalissa tekoälyn (AI) käyttö on vasta alkuvaiheessa mutta laajenee vähitellen kokeiluhankkeiden ja digitalisointipyrkimysten kautta. Esimerkiksi kansallisessa julkisten hankintojen portaalissa on lueteltu 119 tekoölyyn liittyvää sopimusta vuodesta 2009 lähtien, ja lähes puolet niistä on tehty vuosina 2023 ja 2024.³⁸ Tekoölyä sovelletaan esimerkiksi oikeus- ja terveydenhuoltoalalla, ja useat julkiset palvelut ovat ottaneet käyttöön tekoölyä hyödyntäviä chattibotteja, jotka tarjoavat virtuaalista apua palveluissa.

Oikeudellinen ja institutionaalinen kehys

EU:n tekoölyasetus yhdessä eurooppalaisen ja portugalilaisen syrjinnän vastaisen ja tietosuojalainsäädännön kanssa muodostaa oikeudellisen kehyksen algoritmiseen syrjintään puuttumiselle. EU:n säännösten lisäksi lisämahdollisuuksia algoritmisen syrjinnän ehkäisemiseen ja siihen puuttumiseen tarjoaa syyskuussa 2024

37 Tämä on luottamuksellinen raportti, jossa tarkastellaan tekoälyn ja automaattisten päätöksen-tekojärjestelmien käyttöä Portugalin julkisella sektorilla, arvioidaan tekoölyyn liittyviä oikeudellisia ja poliittisia puutteita tasa-arvon ja yhdenvertaisuuden näkökulmasta sekä annetaan suosituksia viranomaisille ja päätöksentekijöille.

38. Portal Base, saatavilla osoitteessa www.base.gov.pt, viitattu 10. marraskuuta 2025.

allekirjoitettavaksi tullut tekoälyä ja ihmisoikeuksia, demokratiaa ja oikeusvaltioperiaatetta koskeva Euroopan neuvoston puiteyleissopimus, jonka tasa-arvoa ja syrjimättömyyttä koskevassa 10 artiklassa kielletään syrjintä ja edellytetään, että kukin sopimuspuoli "sitoutuu ottamaan käyttöön tai pitämään voimassa toimenpiteet, joilla pyritään poistamaan eriarvoisuutta ja saavuttamaan oikeudenmukaisia ja tasapuolisia tuloksia".

Portugalissa oli vuonna 2021 merkittävä algoritmiseen päätöksentekoon liittyvä oikeustapaus, kun Portugalin kansallinen lentoyhtiö TAP käytti algoritmia valitessaan jopa 500 työntekijää irtisanottavaksi tuottavuuden, poissaolojen, kokemuksen, työpanoksen, kustannusten ja pätevyyden kaltaisten parametrien perusteella. Työntekijät kritisoivat avoimuuden puutetta, koska kriteerien painotusta ei julkistettu. Vuonna 2023 Lissabonin työoikeudellisten asioiden alioikeus totesi 23 työntekijän tapauksessa, että joukkoyhtymä oli lainvastainen.

Portugalin hallitus on hyväksynyt useita strategisia välineitä digitalisaation ja tekoälyn vastuullisen käytön edistämiseksi, kuten kansallisen tekoälystrategian (National Artificial Intelligence Strategy)³⁹, joka julkaistiin vuonna 2019 osana kansallista digitaalista osaamista koskevaa aloitetta (National Digital Competences Initiative, INCoDe.2030), ja vuoteen 2030 ulottuvan kansallisen digitaalisen strategian (National Digital Strategy)⁴⁰. Nykytilaa ja puutteita analysoivassa raportissa todetaan kuitenkin, että vuosien varrella lisääntyneet strategiat, suunnitelmat ja ohjelmat voivat olla jopa haitallisia, jos niitä ei koordinoita ja integroida asianmukaisesti.

Portugali perusti vuonna 2024 julkishallinnon digitaalisuutta käsittelevän neuvoston. Siihen sisältyy erityinen tekoälyä käsittelevä tekninen työryhmä, joka valvoo tekoälyasetuksen täytäntöönpanoa. Keskeisessä asemassa on myös aiemmin hallinnon uudistamisesta vastannut virasto, joka muutettiin vuonna 2025 valtion teknologiauudistuksesta vastaavaksi virastoksi. Sen tarkoituksena on varmistaa digitaalisten ja tekoälyjärjestelmien vastuullinen käyttö julkishallinnossa. Julkisten yksiköiden, jotka aikovat hankkia tämän alan tavaroita tai palveluja, kuten tekoälyjärjestelmiä, on ilmoitettava asiasta virastolle, ja virasto voi sitten arvioida näitä hankintoja. Virasto on laatinut oppaan tekoälystä julkishallinnossa sekä riskinarviointivälineen, joka kattaa tietyt etiikkaa, yhdenvertaisuutta ja osallisuutta koskevat kysymykset.

Oikeudellisten ja poliittisten kehysten puutteet

Rajallinen avoimuus ja kansallisen tekoälyn ja automaattisen päätöksenteon järjestelmien rekisterin puute

Portugalin julkishallinnossa käytetyistä tekoälyn tai automaattisen päätöksenteon (ADM) järjestelmistä ei ole kattavaa julkista rekisteriä. Tällä hetkellä mikään julkinen mekanismi ei anna kansalaisille tai valvontaelimille mahdollisuutta tunnistaa, mitä

39. National Artificial Intelligence Strategy, saatavilla osoitteessa www.incode2030.gov.pt/en/estrategia-nacional-de-inteligencia-artificial-en, viitattu 10. marraskuuta 2025.

40. Portugal Digital Strategy – The key to simplification, saatavilla osoitteessa <https://digital.gov.pt/documentos/portugal-digital-strategy>, viitattu 10. marraskuuta 2025.

tekoäly- tai ADM-järjestelmiä julkiset yksiköt käyttävät. Avoimuuden puute voi vaikeuttaa tällaisten järjestelmien mahdollisten syrjintäriskien ja haittojen arviointia.

Perusoikeus- ja yhdenvertaisuuselinten jättäminen tekoälyasetuksen hallintokehyksen ulkopuolelle

Portugalin luettelo tekoälyasetuksen 77 artiklan mukaisesti nimetyistä viranomaisista sisältää yhteensä 14 yksikköä: talous-, puolustus- ja oikeusasioista sekä sisäisestä hallinnosta että koulutuksesta vastaavat viranomaiset ja eri sektorien, kuten tietoliikenteen, elintarviketurvallisuuden, taloudellisen turvallisuuden ja työolojen, valvontaviranomaiset. Perusoikeusinstituutioita, yhdenvertaisuuselimiä tai oikeusasiamiestä ei mainita 77 artiklan mukaisina yksikköinä, ”jotka valvovat tai panevat täytäntöön perusoikeuksien, mukaan lukien oikeus syrjimättömyyteen, suojelua”. Tällaisten elinten puuttuminen luettelosta herättää huolta siitä, miten mahdollisia algoritmisia syrjintätapauksia käsitellään EU:n ja kansallisen lainsäädännön mukaisesti ja miten syrjinnän kohteeksi joutuneet henkilöt voivat johdonmukaisesti käyttää oikeussuojakeinoja tekoälyasetuksen hallintokehyksen puitteissa.

Pirstaloitunut valvonta ja koordinointimekanismit

Vastuu tekoälyn hallinnoinnista on hajautettu useille eri viranomaisille, eikä näiden yksiköiden ja ihmisoikeus- tai yhdenvertaisuusvaltuudet omaavien instituutioiden, kuten oikeusasiamiehen tai Portugalin kansalaisuuksien ja sukupuolten yhdenvertaisuutta käsittelevän komission (CIG) kaltaisten yhdenvertaisuuselinten, välillä ole jäsenneltyjä viestintäkanavia. Yhteistyömekanismin puuttuminen vaikeuttaa mahdollisten valvonnan päällekkäisyyksien tai puutteiden tunnistamista ja korjaamista erityisesti perusoikeuksia ja syrjimättömyyttä koskevien velvoitteiden osalta.

Tietoisuuden, resurssien ja teknisen kapasiteetin rajallisuus

Portugalissa yhdenvertaisuus- ja ihmisoikeusnäkökohtia ei ole vielä järjestelmällisesti integroitu tekoälyn suunnittelu- tai hankintaprosesseihin tai tekoälyn käyttöön julkishallinnossa. Yhdenvertaisuuselimillä ja muilla toimivaltaisilla instituutioilla on rajalliset mahdollisuudet saada käyttöönsä riittävästi tekoälyn teknistä asiantuntemusta, eikä niillä ole riittävästi resursseja, joita tarvitaan tekoälyjärjestelmien teknisiin arviointeihin syrjinnän mahdollisten uhrien tukemiseksi. Portugalin kansalaisuuksien ja sukupuolten yhdenvertaisuutta käsittelevällä komissiolla ja muilla asianomaisilla elimillä on resurssipula, mikä rajoittaa niiden kykyä vastata uusiin haasteisiin, sekä rajalliset valtuudet (esimerkiksi tutkinta- ja riita-asioiden käsittelyvaltuuksien puute), jotka rajoittavat niiden kykyä puuttua algoritmiseen syrjintään liittyviin kysymyksiin.

Keskeiset suositukset

- ▶ **Viranomaisten käyttämien tekoäly- ja ADM-järjestelmien julkisen rekisterin perustaminen:** jos viranomaisille asetetaan velvoite dokumentoida tekoälyjärjestelmät keskitettyyn rekisteriin ja ilmoittaa niiden käytöstä, avoimuutta voidaan parantaa, valvonta- ja yhdenvertaisuuselimien tukea ja kansalaisten luottamusta lisätä.
- ▶ **Yhdenvertaisuuselinten tutkinta- ja täytäntöönpanovaltuuksien tehostaminen sekä perusoikeus- ja yhdenvertaisuuselinten nimittäminen 77 artiklan mukaisiksi kansallisiksi elimiksi:** kun huomioidaan lisäksi yhdenvertaisuuselimien koskevien standardien EU-direktiivien saattaminen osaksi kansallista lainsäädäntöä, tämä vahvistaisi näitä instituutioita niin, että ne pystyisivät tekemään ennakoivaa seurantaa, osallistumaan tekoälyä koskevaan hallintoon ja panemaan tehokkaasti täytäntöön sekä EU:n että kansallisen lainsäädännön mukaisia syrjinnän vastaisia velvoitteita.
- ▶ **Institutionaalisten toimeksiantojen selkeyttäminen ja viranomaisten välisen koordinoinnin vahvistaminen ennen perusoikeus- ja yhdenvertaisuuselinten nimittämistä 77 artiklan mukaisiksi elimiksi:** selkeiden vastuualueiden ja virallisten yhteistyömekanismien määrittely yhdenvertaisuus-, tietosuoja- ja kuluttajansuojaviranomaisten välillä auttaa varmistamaan johdonmukaisen tiedonvaihdon ja tekoälyasetuksen täytäntöönpanon.
- ▶ **Yhdenmukaisten yhdenvertaisuus-, sukupuolten tasa-arvo- ja perusoikeusvaikutusten arviointikehysten luonti kansallisten tekoälystrategioiden, hallinnon uudistamisesta vastaavan viraston oppaan sekä tekoälyasetuksen pohjalta:** näillä kehyksillä voidaan kehittää standardoitu kansallinen lähestymistapa, joka ohjaa viranomaisia algoritmisen syrjinnän riskien arvioinnissa ja auttaa varmistamaan suojatoimien johdonmukaisen soveltamisen julkisella sektorilla.
- ▶ **Perusoikeuksia koskevan koulutuksen ja valmiuksien kehittämisohjelmien tarjoaminen virkamiehille ja tekoälyasetuksen mukaisille toimivaltaisille viranomaisille:** viranomaisille on suositeltavaa tarjota kohdennettua koulutusta ja valmiuksien kehittämismahdollisuuksia, jotta algoritmiseen syrjintään, mukaan lukien sukupuoleen perustuvaan syrjintään, voidaan puuttua.
- ▶ **Syrjimättömyyden ja sukupuolten tasa-arvon periaatteiden sisällyttäminen kattavammin tekoäly- ja ADM-järjestelmien julkiseen hankintaan:** tähän sisältyy tekoälyjärjestelmien valvonnan jatkaminen säännöllisin tarkastuksin, kun järjestelmiä otetaan käyttöön ja käytetään.
- ▶ **Tekoälylukutaidon edistäminen ja tiedotuskampanjoiden järjestäminen:** On suositeltavaa laatia tiedotuskampanjoita ja opetusmateriaaleja, joilla yleisölle tiedotetaan tekoälyn käytöstä julkishallinnossa, sen mahdollisista vaikutuksista ja käytettävissä olevista oikeussuojakeinoista. Tekoälylukutaidon vahvistaminen auttaa lisäämään avoimuutta ja vastuullisuutta näiden järjestelmien käytössä.

- ▶ **Yhdenvertaisuuselinten riittävän rahoituksen ja tekoälyä koskevan teknisen asiantuntemuksen varmistaminen:** yhdenvertaisuuselimille on annettava riittävästi rahoitusta ja tekoälyä koskevaa teknistä erityisasiantuntemusta, jotta ne voivat puuttua algoritmiseen syrjintään.
- ▶ **Sen selvittäminen, kuinka hallituksen meneillään oleva hanke, jossa kehitetään uutta monikanavaisten ihmiskeskeisten julkisten palveluiden sukupolvea gov.pt-alustalla, voi auttaa kansalaisia tekemään tekoälyyn liittyvää syrjintää koskevia valituksia tai käsittelypyyntöjä.**

Lähdeviitteet

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-add1rev, 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,> 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-fight/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-lindence-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.

Selbts 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.

Tekoälyn kasvava käyttö ja integrointi moniin sovelluksiin useilla eri aloilla asettaa merkittäviä haasteita perusoikeuksien suojelulle. Näistä algoritmisen syrjinnän ongelma on erityisen ajankohtainen huolenaihe. Empiiriset tutkimukset ovat osoittaneet, että algoritmiset vinoumat eivät ainoastaan heijasta olemassa olevia yhteiskunnallisia eriarvoisuuksia, vaan ne voivat myös pahentaa niitä, erityisesti aloilla, joilla algoritmisen päätöksenteon käyttö on jo yleistä, kuten poliisitoiminnassa, työelämässä, koulutuksessa ja vakuutus-alalla. Vuonna 2024 sekä Euroopan unioni että Euroopan neuvosto hyväksyivät merkittävät säädökset: EU:n tekoälyasetuksen (asetus 2024/1689) ja Euroopan neuvoston puiteyleissopimuksen tekoälystä ja ihmisoikeuksista, demokratiasta ja oikeusvaltiosta. Tässä raportissa tarkastellaan, miten nämä uudet säädökset vahvistavat suojaa algoritmista syrjintää vastaan Euroopassa, ja arvioidaan oikeudellisten kehyksien puutteita. Raportin tavoitteena on myös seurata ja koota yhteen viimeisimmät kehityssuuntaukset, jotta eri toimijoilla, kuten tasa-arvo- ja yhdenvertaisuuselimillä on riittävät tiedot ja välineet sopeutua muuttuvaan oikeudelliseen tilanteeseen ja tarvittaessa puuttua meneillään oleviin muutoksiin.



YHDENVERTAISUUSVALTUUTETTU
NON-DISCRIMINATION OMBUDSMAN



Euroopan unionin jäsenvaltiot ovat päättäneet yhdistää osaamisensa, voimavaransa ja kohtalonsa. Yhdessä he ovat luoneet vakauden, demokratian ja kestäväen kehityksen alueen, jossa pidetään yllä kulttuurista monimuotoisuutta, suvaitsevaisuutta ja yksilön vapauksia. Euroopan unioni on sitoutunut jakamaan saavutuksensa ja arvonsa muiden maiden ja kansojen kanssa.

www.europa.eu

Euroopan neuvosto on maanosansa johtava ihmisoikeusjärjestö. Sillä on 46 jäsenvaltiota, mukaan lukien kaikki Euroopan unionin jäsenet. Kaikki Euroopan neuvoston jäsenvaltiot ovat allekirjoittaneet ihmisoikeuksien, demokratian ja oikeusvaltion suojaamiseksi tehdyn Euroopan ihmisoikeussopimuksen. Euroopan ihmisoikeustuomioistuin valvoo ihmisoikeussopimuksen täytäntöönpanoa jäsenvaltioissa.

www.coe.int



EUROOPAN UNIONI

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE