HUMAN RIGHTS WATCH

# Q&A: How the EU's Flawed Artificial Intelligence Regulation Endangers the Social Safety Net

*Contents*

# Summary

In April 2021, the European Commission released its 108-page proposal to regulate artificial intelligence ("AI"), describing it as an attempt to ensure a "well-functioning internal market for artificial intelligence systems" that is based on "EU values and fundamental rights."[1] It is the bloc's first major attempt to comprehensively regulate such systems, and could have global repercussions.

The proposed Artificial Intelligence Act comes amid growing concern, in Europe and beyond, about how AI and other forms of algorithmic decision-making are increasingly woven into the social safety net, affecting social and economic rights. The former UN special rapporteur on extreme poverty and human rights, and academics and civil society groups, have warned against the use of high-tech tools to deny or limit access to lifesaving benefits and other social services, at a time when these programs are a critical bulwark against growing poverty and income inequality.[2]

While the EU regulation broadly acknowledges these risks, it does not meaningfully protect people's rights to social security and an adequate standard of living. In particular, its **narrow safeguards neglect how existing inequities and failures to adequately protect rights** – such as the digital divide, social security cuts, and discrimination in the labor market – **shape the design of automated systems, and become embedded by them**.[3]

This Q&A highlights how algorithmic decision-making is transforming the administration of welfare benefits and other social protection measures in Europe, and the human rights concerns created by this transition to algorithmic governance. It also examines how the proposed regulation fails to address these concerns, and what the European Parliament should do to fix these shortcomings.

---

[1] Proposal for a Regulation of the European Parliament and Of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final, April 21, 2021, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 ("Proposed AI regulation").

[2] UN Human Rights Council, Report of the Special Rapporteur on Extreme Poverty and Human Rights, Philip Alston, "on the digital welfare state," A/74/48037, October 11, 2019, https://www.ohchr.org/Documents/Issues/Poverty/A_74_48037_AdvanceUneditedVersion.docx; Lina Dencik and Anne Kaun, "Datafication and the Welfare State: An Introduction," *Global Perspectives*, 2020, https://www.diva-portal.org/smash/get/diva2:1448242/FULLTEXT02.pdf; Human Rights Watch, "UN: Protect Rights in Welfare Systems' Tech Overhaul," October 17, 2019, https://www.hrw.org/news/2019/10/17/un-protect-rights-welfare-systems-tech-overhaul.

[3] See also Agathe Balayn and Seda Gürses, "Beyond Debiasing: Regulating AI and its inequalities," September 2021, https://edri.org/wp-content/uploads/2021/09/EDRi_Beyond-Debiasing-Report_Online.pdf,.

# PART I: Algorithmic Decision-Making and the Right to Social Security

## 1. How does the regulation define "artificial intelligence"?

The definition of artificial intelligence in the regulation covers a wide range of computational methods designed to perform certain tasks, such as generating content, assisting in decision-making about people's social security benefits, predicting an individual's risk of committing fraud or defaulting on a loan, or recommending advertisements on social media platforms.[4]

The regulation is not only concerned with the latest advances in artificial intelligence that have attracted intense government and public scrutiny in recent years, such as the use of self-learning technologies to recognize faces or generate text. It would also cover rules-based algorithms that, by most technical definitions, would not be considered artificial intelligence – in other words, algorithms programmed with a well-defined sequence of rules to make a decision, prediction, or recommendation, such as the amount an individual should receive in their social security payment, the likelihood that they will find a job in the next year, or their grades on a particular exam.[5]

## 2. Do EU member states have a duty to protect the right to social security?

Yes. Article 34 of the Charter of the Fundamental Rights of the European Union "recognises and respects the entitlement to social security benefits and social services" in a broad range of cases, such as when people lose employment.[6] Article 34 further establishes the "right to social and housing assistance so as to ensure a decent existence for all those who lack sufficient resources."

Member states also have a duty to protect the right to social security under international human rights law and regional standards, which may be broader than their corresponding duty under the EU charter.[7]

---

[4] Proposed AI regulation, Art. 3(1), Annex I.

[5] Human Rights Watch, "Automated Hardship," September 29, 2020, p. 16 – 17, https://www.hrw.org/report/2020/09/29/automated-hardship/how-tech-driven-overhaul-uks-social-security-system-worsens; Matt Burgess, "The lessons we all must learn from the A-levels algorithm debacle," *WIRED*, August 20, 2020, https://www.wired.co.uk/article/gcse-results-alevels-algorithm-explained.

[6] Charter of Fundamental Rights of the European Union, 2012/C 326/02, https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012P/TXT&from=EN.

[7] Human Rights Watch, "Protecting Economic and Social Rights During and Post-Covid-19," June 29, 2020, https://www.hrw.org/news/2020/06/29/protecting-economic-and-social-rights-during-and-post-covid-19.

Article 9 of the International Covenant on Economic, Social and Cultural Rights (ICESCR) recognizes the right of everyone to "social security, including social insurance."[8] All EU member states have ratified the covenant.

The UN body that monitors compliance with ICESCR has found that states violate social security rights when they fail to take "sufficient and appropriate action" to protect people's access to benefits (such as by passing weak regulations).[9] The treaty body also warned against adopting "deliberatively retrogressive measures" that deprive people of benefits based on irrational or discriminatory eligibility criteria, such as their place of residence.[10]

International human rights law also recognizes the right of everyone to an "adequate standard of living … including food, clothing, and housing," which often requires securing social security benefits.[11]

All EU member states are members of the Council of Europe and have ratified at least one of the two versions of the European Social Charter (ESC): the 1961 original, and the 1996 revised version.[12] Articles 12, 13, and 14 of the Charter guarantee the right to social security, the right to social and medical assistance, and the right to benefit from social welfare services respectively. Article 30 of the 1996 revised version also guarantees the right to protection against poverty and social exclusion.

In March 2021, the European Committee of Social Rights, which oversees compliance with the Charters, issued a statement on social rights observing "that the rapid digitalisation of social and other services during the pandemic has tended to accentuate the social exclusion of people living in poverty due to lack of equipment, broadband connections and digital skills (the digital divide)."[13]

---

[8] International Covenant on Economic, Social and Cultural Rights (ICESCR), Art. 22, https://www.ohchr.org/EN/ProfessionalInterest/Pages/CESCR.aspx

[9] UN Committee on Economic, Social and Cultural Rights, "The Right to Social Security, Art. 9," General Comment No. 19, E/C.12/GC/19 (2008), https://www.refworld.org/docid/47b17b5b39c.html, para. 64.

[10] Id., para. 65.

[11] ICESCR, Art. 11.

[12] Council of Europe, European Social Charter, European Treaty Series No. 35, https://rm.coe.int/168006b642; Council of Europe, European Social Charter (Revised), European Treaty Series No. 163, https://rm.coe.int/168007cf93.

[13] European Committee of Social Rights, "Statement on Covid-19 and social rights," March 24, 2021, https://rm.coe.int/statement-of-the-ecsr-on-covid-19-and-social-rights/1680a230ca.

## 3. How are governments using automation to monitor and control access to benefits and social security programs?

In Europe, governments have relied on algorithmic decision-making at three key stages in the social security delivery cycle: 1) verifying a person's identity; 2) assessing a person's eligibility for benefits and the amount to which they are entitled; and 3) preventing and investigating benefits fraud. Automating these decisions can make social security administration more efficient for the government. But it can also compromise the privacy of people in need, make it more onerous for them to prove they qualify for government assistance, and profile them in harmful ways.

### *I. Identity verification*

In a bid to curb identity fraud in cash assistance programs, public authorities are capitalizing on technological advances in identity authentication to ensure that benefit applicants are who they say they are. But automating identity checks imposes heightened evidentiary requirements that are intrusive and unduly burdensome for many people who rely on benefits.

In **Ireland**, the Department of Employment Affairs and Social Protection (DEASP), the national welfare office, generally requires applicants for welfare benefits to verify their identity using the Public Services Card (PSC), a government-issued identity card.[14] To obtain a PSC, applicants are required to submit evidence of their identity and address (such as their passport and a utility bill), and schedule an in-person registration appointment at their local welfare office.[15] During the appointment, the office will ask the applicant security questions, record their signature, and take their photo. The DEASP uses facial recognition to compare this photo with data drawn from past applicants' photos to ensure that they have not registered more than once or under a different identity.[16]

The Irish Council for Civil Liberties, a human rights organization, has criticized the DEASP for collecting more personal data than necessary to perform identity checks.[17] It is unclear, for example, why the DEASP collects and analyzes facial images when less invasive means of

---

[14] Social Welfare and Pensions Act 2012, Section 15, https://www.irishstatutebook.ie/eli/2012/act/12/section/15/enacted/en/html#sec15.

[15] Government of Ireland, "Public Services Card: How to apply?," https://psc.gov.ie/how-to-apply/.

[16] Government of Ireland, "UN Questionnaire on the human rights impacts, especially on those living in poverty, of the introduction of digital technologies in the implementation of national social protection systems: Ireland's Response," May 2019, https://www.ohchr.org/Documents/Issues/Poverty/DigitalTechnology/Ireland.pdf.

[17] Irish Council for Civil Liberties, "The Public Services Card," https://www.iccl.ie/2019/the-public-services-card-contd/; Digital Rights Ireland, "Cloak-and-Dagger and the Pandemic Unemployment Payment," July 30, 2020, https://www.digitalrights.ie/cloak-and-dagger-deasp/.

checking people's identity, such as authenticating their passport and proof of address, should ordinarily suffice.[18]

Conditioning access to benefits on facial recognition checks also heightens the risk of discrimination. A 2019 study by the US National Institute for Science and Technology (NIST), a United States government laboratory that has conducted one of the most comprehensive global studies of facial recognition algorithms, has found that facial identification is less accurate for darker-skinned women than for white men.[19] When government agencies use this technology to conduct law enforcement investigations, NIST has warned that these inaccuracies could lead to false accusations.[20] DEASP has not published any information about the accuracy rates of its software.

The former UN special rapporteur on extreme poverty and human rights has also raised concerns that the identity verification process is invasive and exclusionary, because its requirements create high barriers to access benefits.[21] Several PSC applicants who were adopted as children told the online publication *The Journal.ie* that they struggled to provide evidence of their adoption status that was satisfactory to DEASP.[22] Migrants with incomplete identity documents, and rural residents who live far away from the nearest welfare office may also face difficulties registering for a PSC.[23]

Despite these concerns, the DEASP has suspended welfare payments to hundreds of people who have failed to register for the PSC, claiming that the "vast majority" of them are no longer in the country.[24] In some cases, however, people who are eligible have been denied access to essential social security support. In 2017, a woman in her 70s was denied her pension for 18 months – about €13,000 total – when she refused to register for a PSC because she did not think it was

---

[18] Irish Council for Civil Liberties, "The Irish PSC: Enforced digital identities for social protection services and beyond," May 24, 2019, https://www.ohchr.org/Documents/Issues/Poverty/DigitalTechnology/IrishCouncilCivilLiberties.pdf, p. 8.

[19] Patrick Grother, et al., "Face Recognition Vendor Test (FRVT), Part 3: Demographic Effects," National Institute of Standard and Technology, US Department of Commerce, December 2019, https://nvlpubs.nist.gov/nistpubs/ir/2019/NIST.IR.8280.pdf.

[20] Id., p 5.

[21] UN Special Rapporteur on extreme poverty and human rights, OL IRL 1/2020, April 14, 2020, https://spcommreports.ohchr.org/TMResultsBase/DownLoadPublicCommunicationFile?gId=25176.

[22] Cianan Brennan, ""It is the worst form of humiliation" – Woman denied Public Services Card due to adopted status," *thejournal.ie*, January 21, 2018, https://www.thejournal.ie/psc-adoption-refusal-adoption-cert-3806998-Jan2018/; Cianan Brennan, "'They knew I wasn't going to go away' - State in second climbdown as adopted man finally receives PSC eight months later," *thejournal.ie*, April 2, 2018, https://www.thejournal.ie/psc-adoption-card-issued-3929475-Apr2018/.

[23] OL IRL 1/2020.

[24] Elaine Edwards, "Over 450 have welfare suspended for not registering for public services card," *The Irish Times,* February 22, 2018, https://www.irishtimes.com/news/politics/over-450-have-welfare-suspended-fornot-registering-for-public-services-card-1.3401945.

mandatory.[25] The DEASP restored her pension after her situation received widespread media coverage. In 2019, the DEASP denied welfare benefits to a teacher who broke her ankle in an accident and was out of work for seven weeks, after she was unable to travel to the welfare office to register for a PSC.[26] She also objected to the PSC for privacy reasons.

Although the **United Kingdom** is no longer part of the EU, its benefits system, Universal Credit, reflects one of the most ambitious efforts to automate an essential public service, and provides critical insight on how welfare automation can affect rights.[27]

In 2016, the government introduced "GOV.UK Verify", a revamped online process for people to verify their identity when applying for Universal Credit, filing taxes, or using other government services.[28] The service is operated by third-party identity providers. Users of this service are required to upload evidence of their identity, such as a driver's license, bank account details, and mobile phone contracts. The providers check this information against data they have collected (such as people's credit histories) and passport and driver's license data held by the government.

The Verify service has proven difficult to use for the majority of people applying for benefits, since many of them lack the online presence or proof of identity needed to complete its verification checks.[29] In 2019, the UK's National Audit Office, which monitors public spending in Parliament, found that only 38 percent of people who applied for Universal Credit were able to successfully verify their identity using the service.[30] Many people submit evidence of their identity at their local unemployment office instead.

---

[25] Elaine Edwards, "Cutting woman's pension over card 'outrageous,' says Age Action," *The Irish Times*, August 22, 2017, https://www.irishtimes.com/news/social-affairs/cutting-woman-s-pension-over-card-outrageous-says-age-action-1.3194543?mode=sample&auth-failed=1&pw-origin=https%3A%2F%2Fwww.irishtimes.com%2Fnews%2Fsocial-affairs%2Fcutting-woman-s-pension-over-card-outrage.

[26] Cianan Brennan, "Teacher denied access to social welfare after breaking her ankle and refusing to get PSC," *Irish Examiner*, November 18, 2019, https://www.irishexaminer.com/news/arid-30964845.html.

[27] Although the UK is not a member of the EU, the regulation will affect how UK providers design automated systems that they want to place on the EU market. The UK also remains a member of the European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENELEC). According to technology policy experts, the CEN and CENELEC will be responsible for developing technical standards that implement key requirements of the EU AI regulation. All CEN and CENELEC members, including the UK, "commit to adopt identically European Standards and withdraw national conflicting standards" – in this case, standards that govern the design and upkeep of automated systems that are provided to the EU market. *See* "CEN and CENELEC General Assemblies agreed new governance arrangements that will provide a category of full membership for British Member BSI," CEN and CENELEC, June 24, 2021, https://www.cencenelec.eu/news-and-events/news/2021/briefnews/2021-06-24-full-membership-bsi/.

[28] "Guidance: Gov.UK Verify Overview," United Kingdom Government Digital Service, June 18, 2020, https://www.gov.uk/government/publications/introducing-govuk-verify/introducing-govuk-verify.

[29] House of Commons Committee of Public Accounts, "Accessing public services through the Government's Verify digital system," May 8, 2019, https://publications.parliament.uk/pa/cm201719/cmselect/cmpubacc/1748/1748.pdf.

[30] Comptroller and Auditor General, "Investigation into Verify" (London: National Audit Office, 2019), https://www.nao.org.uk/wp-content/uploads/2019/03/Investigation-into-verify.pdf.

In March 2020, when the government imposed Covid-19 lockdown restrictions, there were complaints that the Verify service was slow and unusable, forcing the government to revert to an older verification process for the sudden influx of people claiming Universal Credit.[31] The government also gave people the option to verify their identity over the phone.

In July 2021, the government announced plans to develop a new centralized digital identity platform that will enable a single "sign-on" for access to all government services.[32] It is unclear, however, how this platform will be designed to avoid the problems that plagued the Verify service, which was also intended to centralize identity verification checks.

### *II. Eligibility Assessment and Means Testing*

Governments also use automated decision-making to assess whether people satisfy the eligibility criteria of the benefits they are claiming, and adjust the size of their benefit payments based on changes in their income.

In **France**, the Caisse des Allocations Familiale (CAF), the government agency that oversees the administration of social security benefits, uses an automated system to calculate and distribute benefit payments. In June 2021, CAF admitted that recent changes to the system had caused benefit delays and errors for "1 to 2%" of the approximately six million people receiving a housing allowance known as *Aide personnalisée au logement*, or APL.[33] By this estimate, at least 60,000 to 120,000 people have experienced delays or errors.

These problems stem from a software update that CAF introduced in January 2021, to implement policy changes to APL's means-testing formula. The new formula mandated recalculating the benefit every three months based on the income history over the preceding 12 months of the person receiving it.[34]

---

[31] Bryan Glick, "Huge Queues for Universal Credit Online Applications Due to Coronavirus Surge," Computer Weekly, March 24, 2020, https://www.computerweekly.com/news/252480546/Huge-queues-for-Universal-Credit-online-applications-dueto-coronavirus-surge; "Universal Credit Claimants to Verify Identity Throughout Government Gateway," UK Department for Work and Pensions news release, April 16, 2020, https://www.gov.uk/government/news/universal-credit-claimants-to-verify-identity-throughgovernment-gateway.

[32] Martin Taylor, "A single sign-on and digital identity solution for government," *Government Digital Service*, July 13, 2021, https://gds.blog.gov.uk/2021/07/13/a-single-sign-on-and-digital-identity-solution-for-government/.

[33] Faïza Zerouala, "La réforme des APL vire au cauchemar pour les allocataires et ses agents," *Mediapart*, June 19, 2021, https://www.mediapart.fr/journal/france/190621/la-reforme-des-apl-vire-au-cauchemar-pour-les-allocataires-et-ses-agents/prolonger.

[34] Caisse nationale des allocations familiales, "Apl de juillet 2021 : l'anomalie résolue," August 5, 2021, https://www.caf.fr/allocataires/actualites/2021/apl-de-juillet-2021-l-anomalie-resolue.

*Mediapart,* a French investigative news outlet, reported that the update caused delays to people's benefits or saddled them with debt they did not owe.[35] A 52-year-old woman living in Poitiers told *Mediapart* that she did not understand why CAF deducted 110€ from her benefit payment in April. When she logged on to her CAF account after the deduction, her confusion and anxiety grew when she saw that she owed the agency 1,400€ in overpaid benefits. A case worker she contacted found that this was a software error, but could not correct it in the system. To cope with the benefit cut, she borrowed money from her friends and cut back on groceries and food.

In an article for the digital rights NGO *AlgorithmWatch*, Lucie Inland, a journalist, reported that CAF sent her an automated debt notice in March, two months after the software update.[36] The notice informed her that she owed 542€ in overpaid benefits, and that 60€ would be withdrawn from her benefit payments every month. A caseworker she contacted told her that this was a software error, and CAF subsequently withdrew their decision and canceled the debt. Inland described the ordeal as an "emotional rollercoaster."

In the **United Kingdom**, the government is using a flawed algorithm to calculate people's benefit payments under Universal Credit.[37] The algorithm is supposed to adjust how much people are entitled to each month based on changes in their earnings. But the data the government uses to calculate these changes only reflects the wages people receive within a calendar month, and ignores how frequently people are paid. If an individual receives multiple paychecks in one month – common among people in irregular or low-paid jobs – this can cause the algorithm to overestimate earnings and drastically shrink their payment. This is not just a technical error; the government deliberately chose this method of calculation because it was easier to automate, increased administrative efficiency, and reduced operational costs.[38]

---

[35] Faïza Zerouala, "La réforme des APL vire au cauchemar pour les allocataires et ses agents," *Mediapart*, June 19, 2021, https://www.mediapart.fr/journal/france/190621/la-reforme-des-apl-vire-au-cauchemar-pour-les-allocataires-et-ses-agents/prolonger.

[36] Lucie Inland, "How French welfare services are creating 'robo-debt'," April 15, 2021, https://algorithmwatch.org/en/robo-debt-france/.

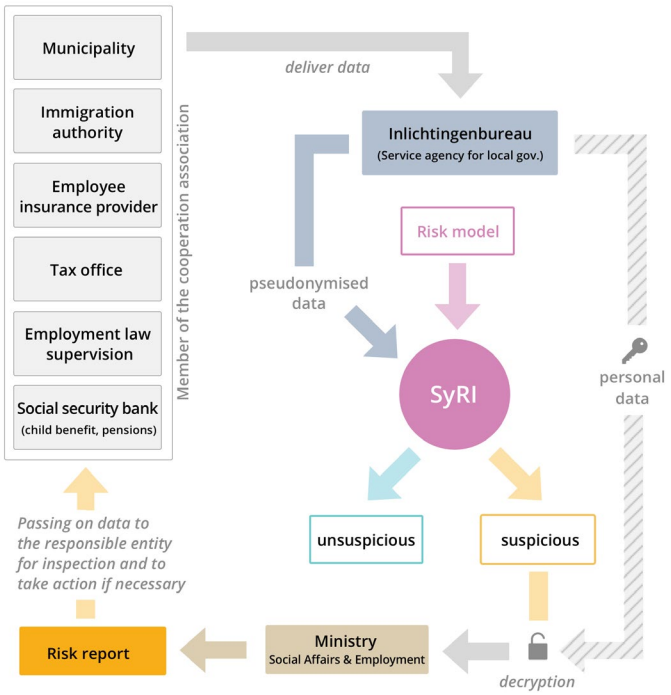
[37] Human Rights Watch, "Automated Hardship," September 29, 2021, https://www.hrw.org/report/2020/09/29/automated-hardship/how-tech-driven-overhaul-uks-social-security-system-worsens.

[38] R (Johnson and others) v SSWP, Court of Appeal (Civil Division), [2020] EWCA Civ 778, June 22, 2020, https://cpag.org.uk/sites/default/files/files/C1.2019.0593-2020-EWCA-Civ-778-R-Johnson-and-others-v-SSWP-FINAL-forHAND-DOWN.pdf, para. 39.

This design choice has led to people going hungry, falling into debt, and experiencing psychological distress. Although an appeals court ordered the government to fix the flaw for a limited number of claimants in June 2020, reports of similar errors persist.[39]

### *III. Countering Fraud*

To help prevent and investigate benefits fraud, some government agencies are using automated decision-making to assign scores to people predicting how likely they are to commit fraud. One of the more prominent examples is *Systeem Risico Indicatie* or SyRI in the **Netherlands**, an automated program developed to score people based on their risk of engaging in benefits and tax fraud.[40] The government experimented with variations of SyRI's technology as early as 2006, and passed a law authorizing the program in 2014.[41] Although the government discontinued SyRI in April 2020, it offers a cautionary tale about the dangers of risk-scoring tools.




*Flowchart showing how SyRI works,* © 2020 AlgorithmWatch.

[39] Child Poverty Action Group, "Early Warning System E-Bulletin – July 2021," July 15, 2021, https://cpag.org.uk/welfare-rights/resources/e-bulletins/early-warning-system-e-bulletin-july-2021.

[40] Amos Toh, "Welfare Surveillance on Trial in the Netherlands," *Open Democracy*, November 8, 2019, https://www.opendemocracy.net/en/can-europe-make-it/welfare-surveillance-trial-netherlands/.

[41] "Subpoena in the main proceedings on substance against the deployment of SyRI by the Department of Social Affairs and Employment," https://pilpnjcm.nl/wp-content/uploads/2019/08/EN-Subpoena-SyRI.pdf, p. 8.

SyRI's secret algorithm assigned risk scores based on an analysis of personal and sensitive data collected by various government agencies, such as employment records, benefits information, personal debt reports, and education and housing histories.[42] When SyRI profiled an individual as a fraud risk, it notified the relevant government agency, which had up to two years to open an investigation into possible criminal behavior or illegal acts.

SyRI, which the government used to profile residents in predominantly low-income neighborhoods, drew widespread criticism for discriminating against poorer people and racial minorities.[43] On February 5, 2020, a Netherlands court held that the lack of transparency about SyRI's risk calculation algorithm made it impossible for people to challenge their risk scores, violating their right to privacy.[44] Two months later, the government announced that it would not appeal the ruling, and halted the use of SyRI.[45] However, civil society groups have raised alarm about a bill in Parliament that would authorize data sharing between public authorities and the private sector to detect fraud and potentially revive the use of risk scoring.[46]

## 4. How else are governments using automation to allocate social security support?

Some governments are using scoring tools to assess the employment prospects of job seekers, and to offer job support measures (such as publicly financed apprenticeships or training courses) based on these predictions. But whether a person will find a job is contingent on complex and highly subjective factors such as their health, educational background, and interpersonal skills. Trying to reduce these factors into a single, individualized score can deny people support they need, and intensify discrimination in the labor market.

In **Austria**, the Public Employment Service (*Arbeitsmarktservice* or "AMS") uses an algorithm to predict a job seeker's employment prospects based on factors such as gender, age group,

---

[42] Amos Toh, "Welfare Surveillance on Trial in the Netherlands," *Open Democracy*, November 8, 2019, https://www.opendemocracy.net/en/can-europe-make-it/welfare-surveillance-trial-netherlands/.

[43] Ilja Braun, "High-Risk Citizens," *AlgorithmWatch*, July 4, 2018, https://algorithmwatch.org/en/high-risk-citizens/.

[44] Amos Toh, "Dutch Ruling a Victory for Rights of the Poor," Human Rights Watch, February 6, 2020, https://www.hrw.org/news/2020/02/06/dutch-ruling-victory-rights-poor.; The Hague District Court, C/09/550982 / HA ZA 18-388, February 5, 2020, https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878.

[45] The Netherlands Ministry of Social Affairs and Employment, "Kamerbrief naar aanleiding van vonnis rechter inzake SyRI," April 23, 2020, https://www.rijksoverheid.nl/ministeries/ministerie-van-sociale-zaken-en-werkgelegenheid/documenten/kamerstukken/2020/04/23/kamerbrief-naar-aanleiding-van-vonnis-rechter-inzake-syri.

[46] Platform bescherming burgerrechten, "SyRI-coalitie aan Eerste Kamer: 'Super SyRI' blauwdruk voor meer toeslagenaffaires," January 11, 2021, https://platformburgerrechten.nl/2021/01/11/syri-coalitie-aan-eerste-kamer-super-syri-blauwdruk-voor-meer-toeslagenaffaires/.

citizenship, health, occupation, and work experience.[47] AMS prioritizes access to its services – such as job search assistance, job placement and training schemes – for job seekers with moderate employment prospects. At the same time, it reduces AMS support to job seekers with low or high employment prospects, reasoning that such support would have a negligible effect on their hiring chances.

The government contends that job seekers with low prospects require different kinds of intervention to "stabilize their personal situation and strengthen their motivation." They are redirected to another government agency that specializes in helping people in "crisis situations" that prevent sustainable employment, such as addiction, debt, and family problems.[48]

Academics and civil society groups have found that the algorithm discounts people's employability when they are women over age 30, women with childcare obligations, are migrants, or have disabilities.[49] Misclassifying people from these groups and limiting them to crisis-oriented support is discriminatory, and could undermine their ability to find and maintain a job and enjoy their right to an adequate standard of living.

The algorithm can also discriminate against people in subtler ways. For example, the algorithm factors the employment rates where the job seeker lives into its prediction of their hiring chances.[50] Since lower employment rates in some areas of the country may be linked to discrimination against racial or ethnic minorities, this seemingly harmless variable can put them at a disadvantage.

In May 2014, **Poland's** Ministry of Labor and Social Policy began using an automated profiling tool that categorized job seekers into three categories: 1) job seekers with appropriate professional qualifications and interpersonal skills; 2) job seekers "requiring support" because their

---

[47] Doris Allhutter, Florian Cech, Fabian Fischer, Gabriel Grill, and Astrid Mager, "Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective," *Frontiers in Big Data*, February 21, 2020, https://www.frontiersin.org/articles/10.3389/fdata.2020.00005/full.

[48] Arbeitsmarktservice, "Beratungs- und Betreuungseinrichtungen (BBE)," https://arbeitplus.at/lexikon/beratungs-und-betreuungseinrichtungen-bbe/.

[49] Allhutter et al., "Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective," https://www.frontiersin.org/articles/10.3389/fdata.2020.00005/full; Paolo Lopez, "Reinforcing Intersectional Inequality via the AMS Algorithm in Austria," Proceedings of the STS Conference Graz 2019, 2020, https://diglib.tugraz.at/download.php?id=5e29a88e0e34f&location=browse; epicenter.works, "Stoppt den AMS-Algorithmus," https://amsalgorithmus.at/.

[50] Doris Allhutter, Astrid Mager, Florian Cech, Fabian Fischer, Gabriel Grill, "Der AMS-Algorithmus: Eine Soziotechnische Analyse des Arbeitsmarktchancen-Assistenz-Systems (AMAS)," Institut für Technikfolgen-Abschätzung der Österreichischen Akademie der Wissenschaften, November 2020, http://epub.oeaw.ac.at/ita/ita-projektberichte/2020-02.pdf, p. 42.

professional skills are limited or outdated; and 3) job seekers that are "distant from the labor market."[51]

The digital rights nongovernmental organization Fundacja Panoptykon found that the government provided scant employment resources to job seekers in the third category, cutting them off from critical job support measures such as publicly financed apprenticeships, childcare reimbursement, and training courses. Job seekers in the third category were disproportionately likely to be women, older people, people with disabilities, and people with little or no formal education. In 2019, the country's Constitutional Tribunal ruled that the government's use of job seekers' personal data to profile them was unconstitutional, because it was done without legislative approval.[52] In December that year, the government discontinued the profiling tool.

[51] Jędrzej Niklas, Karolina Sztandar-Sztanderska, Katarzyna Szymielewicz, "Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making," *Fundacja Panoptykon*, 2015, https://panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf.

[52] Jędrzej Niklas, "Poland: Government to scrap controversial unemployment scoring system," *AlgorithmWatch*, April 16, 2019, https://algorithmwatch.org/en/poland-government-to-scrap-controversial-unemployment-scoring-system/.

# PART II: The EU AI regulation

## 5. How is the regulation's risk-based approach to regulating AI systems expected to work?

The regulation outlines a four-tiered approach to regulating risks posed by AI systems, with different rules applying to each level of risk. The categories are:

- **A ban on "unacceptable risk" (article 5):**[53] The regulation deems certain types of social scoring and biometric surveillance to be an "unacceptable" risk to privacy, nondiscrimination, and related human rights. Public authorities are banned from scoring people's "trustworthiness" in one aspect of their lives (e.g. their ability to repay debt) to justify "detrimental or unfavourable treatment" in another, unrelated context (e.g. denying them the right to travel).[54] Scoring people's trustworthiness to justify "detrimental and unfavourable treatment … that is unjustified and disproportionate to their social behaviour or its gravity" is also prohibited.

  The regulation also imposes a partial ban on law enforcement uses of "'real-time' remote biometric identification systems" in publicly accessible spaces (such as facial recognition cameras in train stations that continuously monitor whether the faces of passers-by match anyone on a watch list).[55] It would, however, permit their use in a troublingly broad range of situations – such as when they are "strictly necessary" to discharge specific law enforcement functions such as finding missing children, preventing terrorist attacks, or identifying the suspect for a criminal offense punishable by a detention sentence of three years or more.[56] It requires judicial authorization for such use, except in emergency situations.

---

[53] Proposed AI regulation, s. 5.2.2 of the Explanatory Memorandum.

[54] Proposed AI regulation, Art. 5(1)(c).

[55] Proposed AI regulation, Art. 5(1)(d).

[56] European Digital Rights, "EU's AI proposal must go further to prevent surveillance and discrimination," April 21, 2021, https://edri.org/our-work/eus-ai-proposal-must-go-further-to-prevent-surveillance-and-discrimination/; Access Now, "EU takes minimal steps to regulate harmful AI systems, must go further to protect fundamental rights," April 21, 2021, https://www.accessnow.org/eu-minimal-steps-to-regulate-harmful-ai-systems/.

- **"High-risk" AI systems (articles 6 and 7)**: The regulation designates an expansive list of AI systems as "high-risk" that would require extra safeguards to deploy.[57] These systems include those used to identify and categorize people based on their biometric data (such as Ireland's facial recognition checks on people applying for identity cards), and manage access to social security benefits and other social services (such as France's housing benefit calculation algorithm).

  Other "high-risk" systems include automated hiring software used to sort through CVs and recruit workers, and law enforcement tools that predict a person's risk of committing a criminal offense. At the heart of regulation is the provision that "high-risk" systems are required to comply with a series of risk mitigation "requirements," which outline the types of information that should be kept and disclosed about a system, safeguards against bias and error, and measures to ensure human oversight, accuracy, robustness, and cybersecurity.[58]

- **Limited-risk AI systems (article 52):**[59] These systems include "biometric categorization," emotion recognition, and deep fake systems.[60] These do not require the same procedures as high-risk systems. The EU's advisory body on data protection has defined biometric categorization as the process of "establishing whether the biometric data of an individual belongs to a group with some predefined characteristic in order to take a specific action" – for example, "an advertising display may show different adverts depending on the individual that is looking at it based on the age or gender."[61]

  Providers and users of limited-risk systems have far fewer obligations than those that develop and sell "high-risk" AI systems. For example, a company that sells intelligent advertising displays is required only to make sure that these displays notify people exposed to them that they are "interacting with an AI system."[62] A department store that uses these displays needs to make sure that customers exposed to them are aware of their

---

[57] Proposed AI regulation, Art. 6, Annex III.

[58] Proposed AI regulation, Chapter 2.

[59] The European Commission describes these systems as "limited risk" systems, but this description is not explicit in the regulation. European Commission, "New rules for Artificial Intelligence – Questions and Answers," April 21, 2021, https://ec.europa.eu/commission/presscorner/detail/en/QANDA_21_1683.

[60] Sam Gregory, "Authoritarian Regimes Could Exploit Cries of 'Deepfake,'" *WIRED*, February 14, 2021, https://www.wired.com/story/opinion-authoritarian-regimes-could-exploit-cries-of-deepfake/.

[61] Article 29 Data Protection Working Party, "Opinion 3/2012 on developments in biometric technologies," 00720/12/EN WP193, April 27, 2012, https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2012/wp193_en.pdf.

[62] Proposed AI regulation, Article 52.

operation. The regulation, however, lets law enforcement off the hook from even these minimal notification requirements when they use this type of system to detect and prevent crime.

- **Minimal-risk AI systems** are all other systems not covered by the regulation's requirements and safeguards.[63]

### 6. Does the regulation prevent government agencies from scoring or profiling people applying for benefits or job support?

It is unclear. The regulation's ban on "trustworthiness" scoring appears to have originally been designed to prevent public authorities from deploying "general purpose" scoring systems – a dystopian future where our lives are reduced to a single social score that controls whether we can board a plane, get a loan, or take certain jobs.[64]

But if the ban is limited to only such "general purpose" scoring systems, then it would fail to capture the reality of current scoring practices or their harm. The fraud detection software previously used in the Netherlands and the employment profiling algorithm in Austria, for example, are more limited in scope. Nevertheless, they can still have devastating human rights consequences, such as when they incorrectly flag people as a fraud risk or deny them the support they need to find a job.

### 7. Does the regulation ban the use of facial recognition to screen people who apply for social security support?

No, since current applications of facial recognition to verify the identity of people who apply for welfare benefits – such as Ireland's registration process for Public Services Cards – do not appear to involve the real-time collection and analysis of facial images in public spaces. Instead, the regulation would classify this use, and most other forms of automated decision-making about social security benefits, as "high-risk" systems subject to heightened safeguards.

---

[63] According to the European Commission, "minimal risk" systems "can be developed and used to subject to the existing legislation without additional legal obligations." European Commission, "New rules for Artificial Intelligence – Questions and Answers," April 21, 2021, https://ec.europa.eu/commission/presscorner/detail/en/QANDA_21_1683.

[64] Proposed AI regulation, Recital 17; Louise Matsakis, "How the West Got China's Social Credit System Wrong," *WIRED*, July 29, 2019, https://www.wired.com/story/china-social-credit-score-system/.

### 8. The bulk of the regulation's safeguards falls on "providers" of welfare technology rather than the agencies that simply use them. Why does this matter?

This unequal distribution of regulatory burden means that harm caused by technologies bought "off-the-shelf" are not regulated as well, even when they are just as severe as those caused by bespoke software.

If a social security agency designs its own software for assessing whether benefit applicants are a fraud risk, they would be a "provider" of a "high-risk" AI system.[65] As a result, the agency is required to comply with risk mitigation requirements, such as disclosing when they began or stopped using the software, and establishing procedures that enable staff to override and correct software errors.

But the same agency would be considered a software "user" rather than a provider if they purchase the risk scoring tool off-the-shelf.[66] This would free them from disclosing their reliance on the tool, and give them broad discretion to outsource their oversight responsibilities to the software provider itself, or a third party.[67]

This discrepancy owes to the regulation's assumption that providers are best placed to forecast, identify, and mitigate the main harms that stem from the systems they create. But this is not always true. In public-private partnerships to deliver social services, for example, public authorities bear ultimate responsibility for ensuring that the software they use protects rights.

### 9. Will the proposed regulation protect people who apply for or receive social security support from algorithmic discrimination?

The regulation mandates providers to examine the data they use to develop "high-risk" systems for biases, gaps, and shortcomings, but these safeguards fail to grapple with the root causes of algorithmic discrimination.

The discriminatory effects of Austria's employment profiling algorithm, for example, are not merely due to the biased inferences it draws from labor market data about the job prospects of women, people with disabilities, and caregivers. The broader purpose of the algorithm is also suspect. According to one study, austerity policies are "one of the main drivers" of the algorithm:[68]

---

[65] Proposed AI regulation, Arts. 3(2), 28(1).

[66] Proposed AI regulation, Art. 3(4).

[67] Proposed AI regulation, Art. 29.

[68] Allhutter et al., "Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective," https://www.frontiersin.org/articles/10.3389/fdata.2020.00005/full; see also Doris Allhutter et. al, "Der AMS-Algorithmus: Eine

> "While the initial goal was to serve a growing number of jobseekers with a stagnating budget, the … algorithm was later seen as a means for budget cuts by the then responsible Ministry of Social Affairs."

The algorithm helps to legitimize those cuts by reinforcing the narrative that people with poor job prospects are "unstable and unmotivated" and undeserving of the state's investments in job seekers. The appearance of mathematical objectivity obscures the messier reality that people's job prospects are shaped by structural factors beyond their control, such as disparate access to education and job opportunities.

Centering the rights of low-income people early in the design process is critical, since correcting human rights harm once a system goes live is exponentially harder. In the UK, the defective algorithm used to calculate people's Universal Credit benefits is still causing people to suffer erratic fluctuations and reductions in their payments, despite a court ruling in 2020 ordering the government to fix some of these errors.[69] The government has also resisted broader changes to the algorithm, arguing that these would be too costly and burdensome to implement.[70]

A more holistic due diligence process should transcend questions of technical bias and examine the human rights implications of automating essential government functions. Algorithmic impact assessments should probe whether a government's attempt to automate access to social security support could result in an overall reduction in people's benefits, or a disproportionate reduction for vulnerable and marginalized groups. These assessments should also examine whether the use of technology imposes more restrictive conditions on access and eligibility (for example, by making it more difficult for people with low digital literacy to provide proof of their identity or submit an online application).

### 10. Will the regulation's transparency requirements help us know more about how governments use AI in social security administration?

The regulation is a step in the right direction, but loopholes are likely to prevent meaningful transparency. The regulation would create a centralized database of high-risk AI systems developed and used in the EU.[71] This is a measure that will help people understand how their government is automating public services. Notably, however, the database only publishes generic

---

Soziotechnische Analyse des Arbeitsmarktchancen-Assistenz-Systems (AMAS)," November 2020, http://epub.oeaw.ac.at/ita/ita-projektberichte/2020-02.pdf.

[69] Child Poverty Action Group, "Early Warning System E-Bulletin – July 2021," July 15, 2021, https://cpag.org.uk/welfare-rights/resources/e-bulletins/early-warning-system-e-bulletin-july-2021.

[70] R (Pantellerisco and others) v SSWP, Court of Appeal (Civil Division), [2021] EWCA Civ 1454, October 10, 2021, https://cpag.org.uk/sites/default/files/CPAG%20Pantellerisco%20v%20SSWP%20CA%20final%20judgment.pdf.

[71] Proposed AI regulation, Art. 60.

details about the status of an automated system, such as whether it's active or discontinued, and in which EU countries the system is active.[72] Disaggregated data critical to the public's understanding of a system's impact, such as the specific government agencies using it, dates of service, and what the system is being used for, will not be available. In other words, the database might tell you that a company in Ireland is selling fraud risk scoring software in France, but not which French agencies or companies are using the software, and how long they have been using it.

Buried in the regulation's dense language are key exemptions from disclosure requirements that could hinder transparency and accountability for AI systems, creating grave risks to rights. Notably, the regulation provides significant exemptions for law enforcement and migration control authorities. For example, providers are supposed to disclose "electronic instructions for use" that explain the underlying logic of how a system operates or makes decisions, such as the source code for the algorithms tasked with calculating benefits.[73] But the regulation states that the information "shall not be provided in the areas of law enforcement and migration, asylum and border control management."[74] As a result, it is likely that critically important information about a broad range of law enforcement technologies that could impact human rights, including criminal risk assessment tools and "crime analytics" software that parse large datasets to detect patterns of suspicious behavior, would remain secret.

There is a risk that this broadly drafted exception will be applied to automated social security systems that also serve a law enforcement purpose. Anti-fraud systems have been designed to flag potential benefits fraud and can trigger criminal investigations. In The Netherlands, immigration authorities, the police, and prosecutors were authorized to request people's SyRI scores in pursuit of an investigation.[75] The UK's Department for Work and Pensions, the country's welfare agency, relies on commercially developed risk scoring software to detect and investigate benefits-related offenses.[76] The regulation could restrict transparency about how tools like these are programmed to make decisions.

---

[72] Proposed AI regulation, Annex VIII.

[73] Proposed AI regulation, Annex VIII(11), Arts. 51 and 60.

[74] Proposed AI regulation, Annex VIII(11), Annex (III)(6).

[75] The Hague District Court, C/09/550982 / HA ZA 18-388, February 5, 2020, https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878, para. 4. 10.

[76] UK Department for Work & Pensions, "Privacy Impact Assessment: Living Together Data & Analytics," March 1, 2017, https://www.whatdotheyknow.com/request/589574/response/1416586/attach/3/Annex%20A%20Privacy%20Impact%20assessment.pdf?cookie_passthrough=1.

### 11. Will the proposed regulation's requirements on human oversight protect people from the failures of welfare automation?

The regulation superficially acknowledges the importance of oversight measures that enable people to understand and override a system's flawed decisions, but fails to set standards that address real-world obstacles to this goal.

The proposal omits a provision from a previous draft that would have allowed people to "decide not to use the high-risk AI system or its outputs in any particular situation without any reason to fear negative consequences."[77] Professor Valerio Di Stefano, an expert on the automation of work, points out that this lack of "explicit workplace protection" could have a chilling effect on human oversight, since workers (such as those working in social security agencies) might hesitate to exercise their formal authority to override a system's outputs without safeguards against retaliation or disciplinary action. [78]

Also missing from the regulation is a deliberative process that involves workers responsible for oversight in the design and implementation of AI systems, as well as training and resource requirements that support them in exercising oversight.[79] These gaps not only create a problematic work environment but also threaten the rights of people they are supposed to safeguard. In France, the *Mediapart* investigation found that caseworkers were experiencing high levels of stress and anxiety because they were struggling to fix software errors that caused erratic reductions and fluctuations in people's benefit payments.[80] These problems mirror those experienced by Universal Credit staff, some of whom told *The Guardian* in 2018 that they were slow or unable to fix social security payment delays or errors that people reported because of confusing or overly prescriptive official guidance.[81]

---

[77] "Draft Regulation on a European Approach for Artificial Intelligence," *Politico,* April 14, 2021, https://www.politico.eu/wp-content/uploads/2021/04/14/AI-Draft.pdf, p. 31.

[78] Valerio De Stefano, "The EU Proposed Regulation on AI: a threat to labour protection?," *Regulating for Globalization*, April 16, 2021, http://regulatingforglobalization.com/2021/04/16/the-eu-proposed-regulation-on-ai-a-threat-to-labour-protection/.

[79] See e.g. Reuben Binns and Valeria Gallo, "Automated Decision Making: the role of meaningful human reviews," April 12, 2019, UK Information Commission Office AI Blog, https://ico.org.uk/about-the-ico/news-and-events/ai-blog-automated-decision-making-the-role-of-meaningful-human-reviews/.

[80] Faïza Zerouala, "La réforme des APL vire au cauchemar pour les allocataires et ses agents," *Mediapart*, June 19, 2021, https://www.mediapart.fr/journal/france/190621/la-reforme-des-apl-vire-au-cauchemar-pour-les-allocataires-et-ses-agents/prolonger.

[81] Patrick Butler, "Universal credit IT system 'broken', whistleblowers say," *The Guardian*, July 22, 2018, https://www.theguardian.com/society/2018/jul/22/universal-credit-it-system-broken-service-centre-whistleblowers-say; James L. Johnson, "I work for the DWP as a universal credit case manager – and what I've seen is shocking," *The Independent*, October 13, 2017, https://www.independent.co.uk/voices/universal-credit-dwp-worker-case-manager-benefits-system-government-food-banks-a7998196.html.

Even though agencies using AI systems serve as a critical backstop against flawed automation, the regulation does not extend oversight requirements to them. The outsourcing of identity verification checks in the United States shows how this loophole could harm human rights. State unemployment agencies across the country are using a commercial platform to verify the identity of people applying for jobless benefits, and relying on the provider to correct verification errors.[82] Many of these agencies do not provide a clear or timely appeals process for benefit errors or denials, or offline alternatives to verify. This creates a single point of failure in the identity verification process, forcing people who are unable to resolve verification issues with the provider to abandon their application altogether. Extending oversight requirements to both providers and the agencies using them gives people directly affected by software errors and other flaws multiple pathways for redress.

### 12. Who will hold providers of welfare technology accountable to the regulation's safeguards?

For **most "high-risk" social security systems**, such as means-testing algorithms and case management software, the regulation allows providers to assess whether their systems are consistent with the regulation's limited rights protections (a process defined as a "conformity assessment procedure based on internal control").[83] Once they sign off on their own systems (by submitting a "declaration of conformity"), they are free to put them on the EU market.[84]

This embrace of self-regulation means that there will be little opportunity for civil society, the general public, and people directly affected by the automation of social security administration to participate in the design and implementation of these systems. The regulation also does not provide people who are denied benefits because of software errors a direct course of action against the provider, or any other means of redress. The government agencies responsible for regulatory compliance in their country could take corrective action against the software or halt its operation, but the regulation does not grant directly affected individuals the right to submit an appeal to these agencies.[85]

For social security and employment systems **with biometric capabilities**, such as anti-fraud checks that require beneficiaries to verify their identity using facial recognition, providers can also sign off on their own systems provided they follow "harmonised standards" that govern the

---

[82] Amos Toh and Lena Simet, "Facial Recognition Problems Denying US Workers Unemployment Lifeline," Human Rights Watch, June 25, 2021, https://www.hrw.org/news/2021/06/25/facial-recognition-problems-denying-us-workers-unemployment-lifeline.

[83] Proposed AI regulation, Art. 43(1)(a), Annex VI.

[84] Proposed AI regulation, Art. 48.

[85] Proposed AI regulation, Arts. 63, 65; Michael Veale, Frederik Zuiderveen Borgesius, "Demystifying the Draft EU Artificial Intelligence Act," *Computer Law Review International* (2021) 22(4), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3896852, p. 24.

technology.[86] These standards translate high-level EU rules that range from safety to accessibility, and environmental impact to concrete technical instructions on how to bring products and services up to code. The harmonized standard on the accessibility of information and communications technology products, for example, requires providers to include design features that are responsive to the needs of users who are deaf and blind.[87]

These harmonized standards are to be developed by privately run nonprofit associations known as European Standardization Organizations.[88] According to EU technology policy experts, the European Committee for Standardization (CEN) and the European Committee for Electrotechnical Standardization (CENELEC) are the most likely candidates to distill the regulation into harmonized standards.[89] In anticipation of this role, both organizations have set up a joint committee dedicated to developing European standards on AI.[90]

The complex process of proposing, amending, and approving standards is generally controlled by CEN and CENELEC members, comprising the national standards bodies of the 27 EU countries, the United Kingdom, Macedonia, Serbia, Turkey, Iceland, Norway, and Switzerland.[91] Civil society participation is largely confined to a trio of European organizations representing consumers, workers, and environmental interests.[92]

Human rights and civil society groups can partner with these organizations to influence the standardization process, but some have raised concern that they have neither the technical know-how nor the resources to participate.[93] It is also unclear whether harmonized standards – which

---

[86] Proposed AI regulation, Art. 43(1).

[87] European Telecommunications Standards Institute, "Accessibility requirements for ICT products and services," EN 301 549 V2.1.2 (2018-08), https://www.etsi.org/deliver/etsi_en/301500_301599/301549/02.01.02_60/en_301549v020102p.pdf.

[88] European Union, Regulation No. 1025/2012, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32012R1025.

[89] Veale and Borgesius, "Demystifying the Draft EU Artificial Intelligence Act," https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3896852, p. 14.

[90] CEN-CENELEC, "CEN-CENELEC JTC 21 'Artificial Intelligence'," https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/.

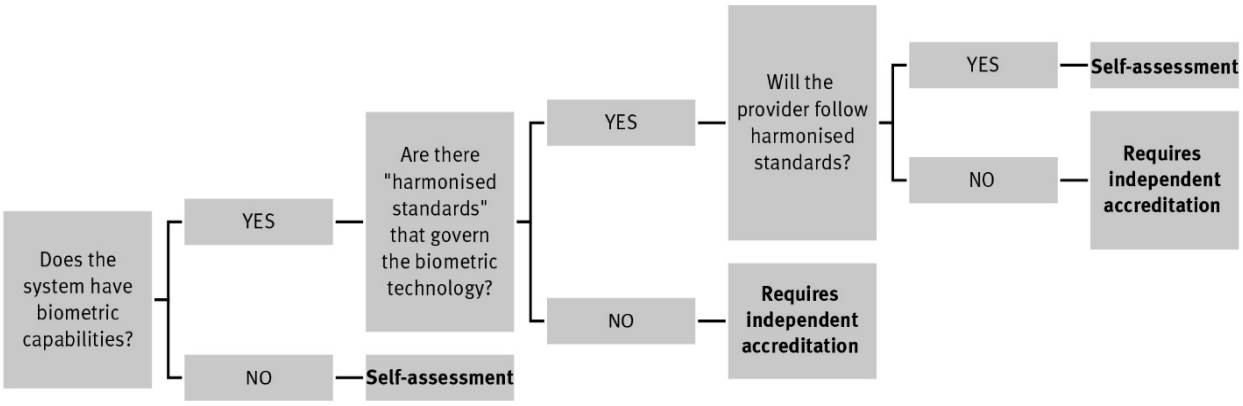
[91] CEN-CENELEC, "European Standards," https://boss.cenelec.eu/homegrowndeliverables/en/pages/eng/.

[92] CEN-CENELEC, "Civil Society: Improving, Strengthening and Legitimising the European Standardization System," https://www.cencenelec.eu/media/CEN-CENELEC/Get%20Involved/Societal%20Stakeholders/civilsocietyleaflet.pdf.

[93] European Center for Not-for-Profit Law, "ECNL Position Paper on the EU AI Act," July 23, 2021, https://ecnl.org/sites/default/files/2021-07/ECNL%20EU%20AI%20Act%20Position%20Paper.pdf, p. 14; European Digital Rights, "European Commission adoption consultation: Artificial Intelligence Act," August 3, 2021, https://edri.org/wp-content/uploads/2021/08/European-Digital-Rights-EDRi-submission-to-European-Commission-adoption-consultation-on-the-Artificial-Intelligence-Act-August-2021.pdf, p. 25; Access Now, "Access Now's submission to the European Commission's adoption consultation on the Artificial Intelligence Act," August 2021, https://www.accessnow.org/cms/assets/uploads/2021/08/Submission-to-the-European-Commissions-Consultation-on-the-Artificial-Intelligence-Act.pdf, p. 30.

are voluntary and non-binding in principle, but could attract widespread adoption in practice – can be challenged in court.

THE EU AI ACT: WHO SIGNS OFF ON AUTOMATED WELFARE SYSTEMS?




### 13. Can a member state establish its own rules to make up for the regulation's shortcomings and protect rights?

Generally, no. The regulation lays down a single set of rules for developing, marketing, and using AI systems; member states cannot impose their own rules "unless explicitly authorized by the Regulation."[94] If a member state requires providers of employment profiling software to disclose which government agencies they are doing business with, the latter could resist this mandate, arguing that it conflicts with the regulation's more limited transparency requirements. Member states may counter that their own rules are needed to protect rights, but the European Court of Justice has set a high bar for such claims.[95] This inflexible design makes it even more critical for the EU to ensure that the regulation incorporates the strongest possible protections for rights.

[94] Proposed AI regulation, Recital 1.

[95] Veale and Borgesius, "Demystifying the Draft EU Artificial Intelligence Act," https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3896852, p. 22 – 23; European Parliament Directorate General for Internal Policies, "EU Social and Labour Rights and EU Internal Market Law," September 2015, https://www.europarl.europa.eu/RegData/etudes/STUD/2015/563457/IPOL_STU(2015)563457_EN.pdf, p. 32 – 37.

# PART III: Recommendations

### 14. How can the EU strengthen its ban on systems that pose "unacceptable risk"?

The starting point of AI regulation should be consistent with relevant human rights standards. In some cases, this will require a clear prohibition on AI applications that threaten rights in ways that cannot be effectively mitigated.

The current proposal to ban some types of "trustworthiness" scoring over a "certain period of time" is vague and impossible to meaningfully implement.[96] Instead, the regulation should prohibit any type of behavioral scoring that unduly restricts or has a negative impact on human rights, including the rights to social security, an adequate standard of living, privacy, and non-discrimination. For example, scoring systems that try to predict whether people are a fraud risk based on records of their past behavior, or serve as a pretext for regressive social security cuts, should be banned.

More broadly, the regulation should codify a strong presumption against the use of algorithms to delay or deny access to benefits. To prevent government agencies from outsourcing harmful scoring practices, the ban should apply to both public authorities and private actors.

### 15. How can the EU ensure that the regulation deals with emerging threats?

The EU should design a system of checks and balances that promptly identifies and addresses human rights concerns with AI applications as they arise. Policymakers are understandably focused on the "unacceptable risks" created by behavioral manipulation, social scoring, and "real-time" biometric identification; systems that the regulation bans in some instances. But this limited catalog of "unacceptable risks" is unlikely to capture every threat that the fast-moving and ever-changing field of AI will pose to human rights.

To ensure that the regulation is responsive to human rights concerns that policymakers cannot foresee, the EU should establish a mechanism for proposing and making additions to the list of systems that pose "unacceptable risk," and a clear and inclusive process for public and civil society participation in such amendments. There should also be a process for upgrading "high-risk" systems to "unacceptable risk" systems, particularly if the deployment of these systems creates unnecessary or disproportionate restrictions on rights.

[96] Proposed AI regulation, Art. 5.

### 16. How can the EU ensure that "high-risk" systems comply with human rights standards?

The regulation should mandate human rights impact assessments throughout the entire life cycle of "high-risk" systems. During the design and procurement phases of an AI system, these assessments could expose human rights risks that may otherwise elude existing risk management processes, such as bias testing or data protection assessments. To ensure that these assessments are meaningful, Human Rights Watch recommends the following amendments:

*I.* *Procurement, Design, and Modification*

- Require users to conduct and publish human rights impact assessments prior to deploying or procuring "high-risk" AI systems, identifying how the system could protect or violate human rights, and the means to prevent or mitigate human rights risks;
- Require these assessments for "substantial modifications" of the AI system that create or amplify human rights risks (for example, modifying the means-testing or risk scoring criteria of an automated system);[97]
- For AI systems that are integrated into welfare and social protection programs, ensure that these assessments incorporate an analysis of the system's benefits and risks to people living in or near the poverty line, such as whether the introduction of the system lowers overall benefit levels or the availability of benefits to informal workers and other groups historically underserved by social security measures;
- Address the obstacles that people with low digital literacy or unreliable internet access might face throughout the life cycle of an AI system, from applying for social security measures to understanding or challenging an automated decision;
- Require government providers and users to provide opportunities to participate in procurement, design, or relevant modification processes, such as through public hearings, notice and comment procedures, and consultations and testing with directly affected stakeholders, such as welfare case workers and beneficiaries; and
- Establish a duty on users to consult relevant public bodies, such as equality bodies and national human rights institutions before deploying "high-risk" AI systems.

*II.* *Post-Deployment*

- Once a system is deployed, require providers to publicly disclose the results of bias audits and subsequent corrective action;
- Require regulatory inspections that assess the human rights impact of "high-risk" systems with reference to their organizational and social context, such as the training, resources,

[97] Article 28(1) of the proposed AI regulation defines "substantial modifications."

and protections provided to workers operating or overseeing these systems, or the socioeconomic factors that make it difficult for people to access, understand, or challenge an AI system;[98]

- Require the creation of an independent oversight body at the national level that is responsible for conducting regulatory inspections;
- Establish a flagging mechanism that gives the public the right to request that the independent oversight body investigate an AI system within its jurisdiction for compliance with human rights standards, including whether human rights impact assessments were adequately performed;
- Adequately staff and resource independent oversight bodies to perform inspections and investigations, ensuring that they have human rights expertise; and
- Require providers or users of systems that consistently fail bias audits or regulatory inspections to cease providing or using these systems.

*III.* *Transparency*

- Clarify that the publication of "electronic instructions for use" of AI systems under Annex VIII requires the publication of all previous and current versions of the source code;
- Remove blanket exemptions to the publication of "electronic instructions for use" in Article 51, Annex VIII, permitting redactions only when necessary and proportionate to a legitimate government aim; and
- Require agencies or other entities using these systems to publish information about their use of "high-risk" AI systems, including their organizational name, the start and end dates of use, and the specific purpose for which they are using these systems (e.g. the specific benefits program for which they are using the system, and the specific tasks assigned to the system).

*IV.* *Oversight*

- Require the development of whistleblower mechanisms and other anti-retaliation safeguards that protect workers when they override or otherwise challenge an automated decision;
- Ensure that public authorities do not leverage the introduction of automated systems to justify reductions in government staff that end up hindering their ability to operate and oversee these systems; and

[98] Ada Lovelace Institute, "Examining the Black Box: Tools for assessing algorithmic systems," April 29, 2020, https://www.adalovelaceinstitute.org/report/examining-the-black-box-tools-for-assessing-algorithmic-systems/, p. 12.

- Require public authorities to provide appropriate training and resources to support staff in operating and overseeing an AI system.

### *V. Remedy and Compliance*

- Establish a right to appeal decisions facilitated by an automated system; and
- Impose fines and other penalties for non-compliance with human rights impact assessments and other risk mitigation requirements, taking into account the nature and severity of the regulatory non-compliance.

*

*Human Rights Watch benefited greatly from expert input provided by Sarah Chander of European Digital Rights (EDRi), Dr. Jedrzej Niklas of Cardiff University, Dr. Michael Veale of University College London, Professor Anne Kaun of Södertörn University, Nina Spurny and Benedikt Gollatz of epicenter.works, Olga Cronin of the Irish Council for Civil Liberties, Professor Tijmen Wisman of Vrije Universiteit Amsterdam, Nicolas Kayser-Bril of AlgorithmWatch and Karolina Iwańska of Fundacja Panoptykon.*