







Legal Definition, AI

EU

Definition on the Proposal for a Regulation on the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts

Article 3.1 - Definitions

"Artificial intelligence system" (AI system) means software that is developed with one or more of the technics and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with".

Annex I Artificial Intelligence Techniques and Approaches

- (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
- (b) Logic and knowledge-beased approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
- (c) Statistical approaches, Bayesian estimation, search and optimization methods.

Example, A				
	🕅 Texto 🖪 Imagens 📄 Documentos 💮 Site	s		
	DETECTAR IDIOMA INGLÊS HOLANDÊS FRANCÊS	~ ÷	→ CROATA NORUEGUÊS FRANCÊS ✓	
	Hi. Am I an artificial intelligence system	1? ×	Bok. Jesam li ja sustav umjetne inte	ligencije? 🏠
	•0	43/5.000 📰 *	4)	0 ⁶ 9 <
Google Translate				Enviar feedback
1. Generate content		00		
2. Text generation.		0		
language translation	1	Histórico Salv	vas Contribuir	
(CoE)				













- → Dutch social security fraud detection.
- → The algorithmic system used a risk classification model to estimate the risk of fraud and inaccurate childcare benefit allocated to families in the Netherlands.
- → In the Netherlands, the government reimburses part of the costs spent by parents and caregivers in daycare. The amount is calculated based on the income of the family. Low-income families can have up to 96% of their daycare costs reimbursed, while families with high income may have 33% of the daycare costs reimbursed.
- → In rough terms, the algorithm took into account the nationality of individuals to create risk profiles of individuals receiving the benefit
- → As a consequence, thousands of low and middle-income families with foreign origins were subject to high scrutiny, falsely accused of fraud and required to pay back the benefits that, in many cases, amounted to tens of thousands of Euros.
- → Other consequences for the families: ended-up in more poverty, because they were wrongly asked to pay back large sums of money; some lost their homes; more than 1000 children were placed in state custody as a result of the accusations.
- → Resignation of the Dutch government in 2021.



XENOPHOBIC MACHINES DISCRIMINATION THROUGH UNREGULATED USE OF ALGORITHMS IN THE DUTCH CHILDCARE BENEFITS SCANDAL







Data training

- Labeling concepts can lead to discrimination when the labels somehow correlate with protected classes. E.g. In the Dutch tax authority system, nationality was correlated as a risk factor for social benefits fraud. This put non-Dutch citizens in a disadvantageous position compared to to Dutch nationals.
- **Biased data** used to feed AI systems. E.g COMPAS used to make decisions about pretrial release and sentencing.
- Identification of proxies. When non-protected aspects, such as the address, may serve as a reason for discrimination.





Assessment List on Trustworthy AI (ALTAI)

• Prior to assessing the trustworthiness of the system, a Fundamental Rights Impact Assessment (FRIA) should be performed.

→ "A FRIA could include questions such as the following – drawing on specific articles in the Charter and the European Convention on Human Rights (ECHR) its protocols and the European Social Charter.

1. Does the AI system potentially negatively discriminate against people on the basis of any of the following grounds (non-exhaustively): sex, race, color, ethnic or social origin, genetic features, language, religion or belief or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation? "Have you put in place processes to test and monitor for potential negative discrimination (bias) during the development, deployment and use phases of the AI system?"

The performance of the ALTAI may involve a multidisciplinary team of people: AI designers and developers; data scientists; procurement officers or specialists; front-end staff that will
work with the AI system; legal/compliance officers; management staff.

REQUIREMENT #4 Transparency

1. Traceability: e.g. "Can you trace back which data was used by the AI system to make a certain decision(s) or recommendation(s)?; "Can you trace back which AI model or rules led to the decision(s) or recommendation(s) of the AI system?"

2. Explainability: e.g. "Did you explain the decision(s) of the AI system to the users?"; Do you continuously survey the users if they understand the decision(s) of the AI system?" REQUIREMENT #5 Diversity, Non-discrimination and Fairness

1. Avoidance of Unfair Bias: e.g. "Did you assess and put in place processes to test and monitor for potential biases during the entire lifecycle of the AI system? Did you test for specific target groups or problematic use cases?"

REQUIREMENT #7 Accountability

1. Auditability: e.g "Did you establish mechanisms that facilitate the AI system's auditability (e.g. traceability of the development process, the sourcing of training data and the logging of the AI system's processes, outcomes, positive and negative impact?); Did you ensure that the AI system can be audited by independent third parties?

Source: European Commission – Shaping Europe's Digital Future

17



How to address AI discrimination: Indirect discrimination

Indirect discrimination

"indirect discrimination shall be taken to occur where an apparently neutral provision, criterion or practice would put persons of [a protected ground] at a particular disadvantage compared with other persons, unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary".

Ref. Art 2, 2 (b) Directive 2000/43/EC; Art. 2, 2 (b); Directive 2000/78/EC; Art 2,1(b) Directive 2006/54/EC.

Approach: By placing the AI system as a neutral provision, criterion or practice, the focus shifts from the technical operation of the AI to its impact.

Advantage: Avoid technical explanations on the AI system.

Disadvantage: Justification. There are much more possibilities to justify indirect discrimination than direct discrimination. Indirect discrimination will only have legal consequences if the use of the AI system is not a proportionate means of achieving a legitimate aim. E.g. business necessity.

How to address AI discrimination: Direct discrimination

Direct discrimination

"Direct discrimination shall be taken to occur where one person is treated less favorably than another is, has been or would be treated in a comparable situation on grounds of [a protected aspect]".

Ref. Art 2, 2 (a) 2000/43/EC; Art. 2, 2 (a); Directive 2000/78/EC; Art 2,1(a) Directive 2006/54/EC.

Approach: In the EU, no intention is required for the legal existence of direct discrimination. In this case, if biased data is used to feed an AI system, for instance, and, as a consequence, a person with a protected aspect is less favorably treated on grounds of her/his protected aspect, direct discrimination takes place.

Advantage: Very exceptionally justifiable. E.g. genuine and determining occupational requirement, provided that the requirement is proportionate to the legitimate objective.

